American Society of Human Genetics 71st Annual Meeting
Abstracts

Plenary Platform Poster talks Poster Presentations

PrgmNr 1012 Multi-trait Bayesian genome-wide association study identifies novel endometrial cancer risk loci at 7q22.1, 8q24.3 and 16q22.2

PrgmNr 1013 Novel basal cell carcinoma susceptibility loci identified from a multiethnic genome-wide meta-analysis

PrgmNr 1014 A fine-mapping study of over 254,000 Asian and European descendants identifies putative susceptibility genes for colorectal cancer

PrgmNr 1015 Assessment of genetic susceptibility to multiple primary tumors through whole-exome sequencing in two large cohorts

PrgmNr 1016 Validation of the BOADICEA multifactorial breast cancer risk prediction model in a large prospective cohort study

PrgmNr 1017 High resolution functional mapping credible causal variants to target genes at 54 common melanoma susceptibility loci

PrgmNr 1020 Analysis of PI3K-AKT-MTOR spectrum disorders through deep genomic and functional models reveals new clinical insights and distinct molecular pathomechanisms

PrgmNr 1021 Skewed X-chromosome inactivation in unsolved neurodevelopmental disease cases can guide re-evaluation for X-linked genes

PrgmNr 1022 Understanding trajectories of neuronal differentiation underlying neurodevelopmental disorders using single-cell technologies

PrgmNr 1023 De novo missense and truncating variants in ZMYND8 result in a distinctive neurodevelopmental disorder

PrgmNr 1024 Extremely rare variants in EIF4A2 are associated with a neurodevelopmental disorder characterized by hypotonia, intellectual disability and epilepsy

PrgmNr 1025 Autism spectrum disorder trios from consanguineous populations are enriched for rare biallelic variants

PrgmNr 1028 Long read sequencing to improve diagnostic rate in critically ill patients

PrgmNr 1029 Establishing nonhuman primate models of human disease using a publicly available rhesus macaque genomic variant database

PrgmNr 1030 Utilization of Dual-Label Optical Genome Mapping for genetic/epigenetic diagnosis

PrgmNr 1031 2D and 3D human neuronal cell models of CHD2-associated epilepsy

PrgmNr 1032 Long-read Sequencing Identifies Missing Disease-Causing Variants and Resolves Complex Structural Variants

PrgmNr 1033 Using CRISPR-based saturation genome editing to improve the diagnosis of neuro-developmental disorders

PrgmNr 1036 Single cell “whole embryo” phenotyping of developmental disorders

PrgmNr 1037 Identifying systematic variation at the single-cell level by leveraging low-resolution population-level data

PrgmNr 1038 Discovery of target genes and pathways of blood trait loci using pooled CRISPR screens and single cell RNA sequencing

PrgmNr 1039 SIMBA-single cell embedding along with features

PrgmNr 1040 Deciphering the variation of gene regulation among individuals via integrative single cell multi-omics analysis

PrgmNr 1041 scATAC.Explorer: A Single-Cell ATAC-Seq Database and Search Tool

PrgmNr 1044 A statistical framework to assess replicability of signals from trans-ethnic genome-wide association meta-analysis: Applications to smoking/drinking addiction traits using 3.4 million individuals

PrgmNr 1045 Pleiotropic decomposition regression to characterize multi-trait genetic architecture

PrgmNr 1046 DeepNull: Modeling non-linear covariate effects improves phenotype prediction and association power

PrgmNr 1047 SR-TWAS: Leveraging multiple reference panels to improve TWAS power by ensemble machine learning

PrgmNr 1048 Genetic and environmental correlations between complex phenotypes differ by race/ethnicity and sex

PrgmNr 1049 Enrichment for gene-by-environment candidate SNPs with variance locus analysis: Extension of vQTL to binary traits in the Pan-UK Biobank

PrgmNr 1058 Long-Term systemic expression and cross-correction ability of HMI-203: Investigational gene therapy candidate for mucopolysaccharidosis type II or Hunter Syndrome

PrgmNr 1059 Tasimelteon Safely and Effectively Improves Sleep in Smith Magenis Syndrome: results from a Double-Blind Randomized Trial Followed by an Open-Label Extension

PrgmNr 1060 Unravelling African genomes: Whole-genome sequencing of 1000 Nigerian samples spanning 50 tribal groups provides new insights into diversity and admixture

PrgmNr 1061 NOTCH3 p.Arg1231Cys is Present in 1 in 92 Pakistani and Associated with Stroke

PrgmNr 1062 A high-resolution panel for uncovering repeat expansions that cause ataxias

PrgmNr 1063 Deployment of clinical whole genome sequencing in support of more than 1,000 resource-limited patients: four years of the iHope Program

PrgmNr 1066 Accurate multi-cancer detection and typing by genome-wide cell-free DNA profile data mining

PrgmNr 1067 A novel method identifies liquid biopsy markers based on relative chromatin accessibility

PrgmNr 1068 Distinct intratumoral oncogenic mechanisms between single cells in primary Ductal Carcinoma In Situ exposed by Primary Template-directed Amplification

PrgmNr 1069 Full-length transcriptome analysis of liver cancer using a long-read sequencing technology

PrgmNr 1070 Functional genomics uncover novel regulatory genomic regions associated with gene expression and survival in multiple myeloma

PrgmNr 1071 Genetic map of human brain folding and links to developmental pathways and disease

PrgmNr 1074 Trans-ancestry imputation and exome sequencing of more than 1 million individuals identifies genetic variation protecting against SARS-CoV-2 infection and predicts individuals at risk for severe COVID-19 outcomes

PrgmNr 1075 Rare variant analyses in 239,395 whole exome and whole genome sequenced participants of the UK Biobank reveals novel genetic associations with renal function and chronic kidney disease

PrgmNr 1076 Novel genetic associations for rare diseases with GWAS and trans-ethnic analysis of self-reported medical data

PrgmNr 1077 Common and rare variant analysis of 21K psoriasis cases and 623K controls identifies novel, protective associations in several genes in the type 1 interferon pathway

PrgmNr 1078 Investigating genetic and phenotypic associations for 168 blood metabolites in 120K UK Biobank participants

PrgmNr 1079 Practical implementation of polygenic risk scores and absolute risk score estimation across diverse ancestry groups

PrgmNr 1103 Characterization of molecular and motor phenotypes in CGG knock-in mice with CRISPR mediated deletion of the trinucleotide repeat

PrgmNr 1104 Upregulation of disease compensatory gene via CRISPR activation in muscular dystrophy mice

PrgmNr 1105 Neuronal gene interaction:Scn8a, Kcna1 and Kcnq2

PrgmNr 1106 Adjunct treatment with glycogen synthase (GYS1) antisense oligonucleotides: Reduce glycogen in the Pompe disease mouse model

PrgmNr 1107 Transethnic evaluation of copy number burden in epilepsies from a genome-wide study of 34,132 subjects

PrgmNr 1108 Variant-to-gene-mapping of insomnia GWAS loci followed by a neuron-specific RNAi screen in Drosophila melanogaster implicates PIGQ at the ‘WDR90’ locus as a novel regulator of sleep

PrgmNr 1111 Somatic Mosaicism and Cardiovascular Disease: Uncovering Hidden Genetic Etiology with Clinical Implications

PrgmNr 1112 Identifying the role of alternative splicing during human heart organogenesis

PrgmNr 1113	Genetic deficiency of the AIM2 inflammasome mitigates JAK2p.V617F clonal hematopoiesis associated cardiovascular diseases risk
PrgmNr 1114	FHL5, a Novel Coronary Artery Disease Gene, Regulates Smooth Muscle Cell Contractility
PrgmNr 1115	Identifying causal protein-cardiometabolic trait relationships using whole genome sequencing
PrgmNr 1116	Network medicine-based multimodal omics discovery and iPSC-based validation of metformin for potential treatment of atrial fibrillation
PrgmNr 1119	The impact of evolutionary processes in shaping the genetics of complex traits in East Asia: A specific contribution from Denisovan introgression
PrgmNr 1120	The influence of Neanderthal introgression on human circadian biology
PrgmNr 1121	Systematic genomic investigation finds evidence of germline and somatic roles of Neanderthal introgression in liver cancer
PrgmNr 1122	Biobank-scale inference of ancestral recombination graphs reveals association to rare and common variants in complex traits
PrgmNr 1123	Local ancestry inference for ~7,000 genomes in the Genome Aggregation Database characterizes fine-scale allele frequencies within admixed American populations
PrgmNr 1124	Sex-specific effective population size inference in recent history
PrgmNr 1127	Local ancestry allows for improved genomic prediction in underrepresented and admixed populations
PrgmNr 1128	Polygenic risk prediction of obesity across the life course and in diverse populations
PrgmNr 1129	Improving Polygenic Prediction in Ancestrally Diverse Populations
PrgmNr 1130	A trans-ancestry polygenic test to predict severe hypercholesterolemia in diverse ancestry patients
PrgmNr 1131	Phenome-wide association study of polygenic risk for asthma in the UK Biobank highlights traits with shared genetic architecture and sex specific effects
PrgmNr 1132	Modelling hidden genetic risk from family history for improved polygenic risk prediction
PrgmNr 1135	A genome-wide mutational constraint map quantified from variation in 76,156 human genomes
PrgmNr 1136	Assessment of mitochondrial genome constraint in the gnomAD database of >50,000 individuals using a composite likelihood model
PrgmNr 1137	Quantifying negative selection on synonymous variants
PrgmNr 1138	Objective targeting of de novo loss of function variants in constrained genes improves diagnostic rates in the 100,000 Genomes Project
PrgmNr 1139	Biallelic noncoding regulatory mutations in autism spectrum disorder
PrgmNr 1140	Estimating the clinical diagnostic yield of whole-genome sequencing in 825 orofacial cleft trios
PrgmNr 1143	Modeling tissue co-regulation to quantify tissue-specific contributions to disease heritability
PrgmNr 1144	Uncovering context-specific genetic regulation of gene expression from single-cell RNA-sequencing using latent-factor models
PrgmNr 1145	Acetylated chromatin domains link chromosomal organization to cell and circuit level dysfunction in schizophrenia and bipolar disorder
PrgmNr 1146	Genetic determinants of alternative splicing in stimulated iPSC-derived macrophages enhance the understanding of immune-mediated disease risk
PrgmNr 1147	Leveraging single-cell ATAC-seq to identify disease-critical fetal and adult brain cell types
PrgmNr 1148	Computational and functional characterization of the hs737 enhancer in autism
PrgmNr 1151	CAPRIN1 haploinsufficiency causes a novel neurodevelopmental disorder associated with morphological and functional impairment in hiPSCs-derived cortical neurons
PrgmNr 1152	Biallelic ATG7 variants impair autophagy leading to neurological disease

PrgmNr 1153 --Bi-allelic variants in the autophagy gene ATG4D are associated with a pediatric neurological disorder characterized by hypotonia, joint laxity, and delayed speech and motor development

PrgmNr 1154 Large-scale, multi-ethnic resource of gene, isoform, and splicing regulation in the developing human brain

PrgmNr 1155 Biallelic variants in CELSR3 cause a syndrome comprising CAKUT, anorectal and CNS anomalies

PrgmNr 1156 Statistical framework uncovers deep intronic splice gain variants implicated in undiagnosed cases

PrgmNr 1159 The impact of copy number variants on complex human traits

PrgmNr 1160 PAV: An assembly-based approach for discovering structural variants, indels, and point mutations in long-read phased genomes

PrgmNr 1161 Classes of rare and common variants differentially contribute to variably expressive phenotypes in complex disorders

PrgmNr 1162 Modeling haplotype sharing powers CNV detection and reveals large-effect phenotype associations

PrgmNr 1163 Multi-tool discovery and joint genotyping of structural variations in 138,134 multi-ethnic TOPMed WGS samples

PrgmNr 1164 Sequencing individual genomes with recurrent deletions reveals allelic architecture and disease loci for autosomal recessive traits

PrgmNr 1167 Web-based analysis and collaboration tool, seqr, for rare disease genomics

PrgmNr 1168 Mendelian diseases and their role in a pediatric neurodegeneration cohort

PrgmNr 1169 A genome-first approach to rare variants in hypertrophic cardiomyopathy genes MYBPC3 and MYH7 in a medical biobank

PrgmNr 1170 Incomplete penetrance of disease variants in the UK Biobank

PrgmNr 1171 Biased pathogenic assertions of loss of function variants challenge molecular diagnosis of admixed individuals

PrgmNr 1172 Leveraging human phenotype ontology based similarity to further functional annotation of the human genome

PrgmNr 1175 Attitudes in four Latin American countries towards donation of personal genetic data for research

PrgmNr 1176 Patient completion of updated hereditary cancer genetic testing

PrgmNr 1177 Awareness and Use of Genetic Testing: An Analysis of the Health Information Trends Survey

PrgmNr 1178 Health Locus of Control, Belief in Genetic Transmission of COVID-19, and COVID-related Outcomes in a Global Sample

PrgmNr 1179 Identifying message design considerations to motivate cascade testing for familial hypercholesterolemia: Stakeholders' advice for healthcare providers

PrgmNr 1180 Usual care genetic counseling vs. ARIA: An RCT of an Accessible, Relational, Inclusive and Actionable model for sequencing results disclosure to historically underrepresented populations in the CSER CHARM study

PrgmNr 1183 Identification of clinically relevant variants in homologous regions in 41,755 exomes

PrgmNr 1184 Analytical validation of clinical whole genome sequencing across a diverse mutation spectrum

PrgmNr 1185 Genome sequencing as a first-line genetic testing in ill newborns

PrgmNr 1186 Genome-to-Treatment: A virtual, automated system for population-scale diagnosis and acute management guidance for genetic diseases in 13.5 hours

PrgmNr 1187 Large scale Cas9 mediated depletion of highly abundant transcripts to expand the interpretable genome and improve the diagnostic yield of clinical RNA Seq

PrgmNr 1188 Integration of proteomics with genomics, transcriptomics and phenomics increases the diagnostic rate of Mendelian disorders

PrgmNr 1191 Integrative analysis of human genetic association studies, single cell transcriptomics, and knockout mouse models identifies cell types and genes involved in skeletal disease

PrgmNr 1192 A modular massively parallel reporter assay uncovers context-specific allelic activity of GWAS variants

PrgmNr 1193 Understanding cell-type-specific drug effects while controlling for genotype

PrgmNr 1194 Modeling eQTLs at single-cell resolution identifies >2000 eQTLs with differential effects across continuous T cell states

PrgmNr 1195 Sex differences in the expression and genetic regulation of drug metabolism and transporter genes in human liver

PrgmNr 1196 Identifying disease-critical cell types and cellular processes across the human body by integration of single-cell profiles and human genetics

PrgmNr 1240 Interpreting Genome-Wide Association Studies of Inflammatory Bowel Disease Through the Lens of Single-Cell Sequencing

PrgmNr 1241 Identifying genetic determinants of transcriptional response dynamics at single cell resolution

PrgmNr 1242 Opposing evolutionary pressures drive clonal evolution and health outcomes in the aging blood system

PrgmNr 1243 Trans-ethnic meta-analysis and functional characterization highlights gene regulation underlying the genetic architecture of telomere length

PrgmNr 1244 Genetic associations from the COVID-19 Host Genetics Initiative highlight biology behind

PrgmNr 1245 Whole Genome Sequencing (WGS) based Association Study of 21 Inflammation Biomarkers in up to 38,473 Multi-Ethnic Individuals Identifies Novel Signals

PrgmNr 1248 Potential regulatory effect of the type 2 diabetes associated KCNQ1 locus on CDKN1C expression identified using induced pluripotent stem cell derived pancreatic islet-like cells

PrgmNr 1249 Single-cell gene expression and chromatin accessibility profiling of human pancreatic islets at basal and stimulatory conditions nominates mechanisms of type 1 diabetes genetic risk

PrgmNr 1250 The Kidney genome atlas reveals a novel locus on chromosome 14 associated with adult proteinuric kidney diseases

PrgmNr 1251 Computational and functional gene prioritization from a saturated GWAS of adult height in >5 million people

PrgmNr 1252 Genetic and functional characterisation of the Melanocortin 3 Receptor (MC3R) links nutritional state to growth, body composition and the onset of puberty

PrgmNr 1253 Causal signatures in skeletal muscle single nucleus multi-omics data across 287 individuals

PrgmNr 1256 A global biobank study of asthma identifies novel associations, illuminates shared genetic architecture, and improves polygenic prediction across diverse ancestry groups

PrgmNr 1257 Genome-wide polygenic risk score of prostate cancer in African and European ancestry men

PrgmNr 1258 Genetic association of phenotypes derived by self-supervised deep learning of retina fundus images reveals new genes for eye development

PrgmNr 1259 Fine-mapping across diverse ancestries drives the discovery of putative causal variants underlying human complex traits and diseases

PrgmNr 1260 Analysis across Taiwan Biobank, Biobank Japan and UK Biobank identifies hundreds of novel loci for 36 quantitative traits

PrgmNr 1261 The effects of demographic-based selection bias on GWAS results in the UK Biobank

PrgmNr 1264 HDAC9 structural variants disrupting TWIST1 transcriptional regulation lead to craniofacial and limb malformations

PrgmNr 1265 De-Novo Mutations identified in Nonsyndromic Cleft lip/Palate Families from Africa

PrgmNr 1266 Identification of novel molecular pathways in syndromic orofacial clefting

PrgmNr 1267 RERE deficiencycontributes to the development of orofacial clefts in humans and mice

PrgmNr 1268 Single cell transcriptomics-directed investigation of de novo mutations and rare inherited genetic variants in cleft lip and palate

PrgmNr 1269 A novel 3 MB deletion on 6p24 removes distant neural-crest enhancers controlling TFAP2A resulting in mild Branchiooculofacial syndrome in a multiplex family with orofacial clefting

PrgmNr 1272 Increasing Detection of Familial Hypercholesterolemia in an Academic Medical Center: The Familial Hypercholesterolemia Identification Registry (FHIRE).

PrgmNr 1273 Finding the Sweet Spot: Exploring patients' perspectives on the adoption of AI and Chatbots in genetic service delivery

PrgmNr 1274 A Step Towards Real-World Population Screening for Hereditary Cancer and Hypercholesterolemia

PrgmNr 1275 Using Supervised Machine Learning to Accelerate Variant Curation and Genomic Diagnosis in Suspected Mendelian Disorders

PrgmNr 1276 Design and implementation of genome interpretation for coronary disease - including monogenic and polygenic risk assessment - in a Preventive Genomics Clinic

PrgmNr 1277 Innovative approaches to optimizing the electronic health record for genomic medicine

PrgmNr 1280 Understanding clonal distributions during human cortical development deciphered through brain somatic mosaicism

PrgmNr 1281 MYCN gain-of-function variants induce excess proliferation of neurons and cause a novel megalencephaly syndrome

PrgmNr 1282 Gain and loss of function variants in EZH1 disrupt neuronal differentiation and lead to overlapping neurodevelopmental disorders

PrgmNr 1283 Genomic atlas of proteome from three tissues relevant to neurological disorders prioritizes proteins implicated in brain diseases

PrgmNr 1284 Single-cell RNA-sequencing of human brain regions identifies cell types and states relevant to disease

PrgmNr 1285 Joint transcriptome and chromatin accessibility profiling of the developing human brain cortex at the single cell level

PrgmNr 1335 Genotype first: Clinical genomics research through a reverse phenotyping approach

PrgmNr 1336 The knowledge gap between pharmacogenomics and medical genetics: Is there a bridge in between?

PrgmNr 1337 Population-based penetrance of clinical variants in individuals of diverse ancestries and ages

PrgmNr 1338 Common-variant dysregulation of Mendelian disorder genes contributes to complex disease

PrgmNr 1339 Understanding the importance of racial diversity in the development and implementation of pharmacogenomic guidelines in kidney transplant patients receiving tacrolimus

PrgmNr 1340 Genome-wide association of structural variations with lipid metabolism traits in TOPMed GOLDN study

PrgmNr 1343 Genetic mapping across autoimmune diseases reveals shared associations and mechanisms

PrgmNr 1344 Systematic discovery of gene-environment interactions for metabolic serum biomarkers

PrgmNr 1345 Global Biobank Meta-analysis Initiative: power genetic discovery for human diseases with > 2.6 million samples across diverse ancestries

PrgmNr 1346 Phenome-wide HLA association landscape of 235,000 Finnish biobank participants

PrgmNr 1347 Global discovery of genetic risk variant allelic enhancer activity in allergic eosinophilic esophagitis and atopic dermatitis

PrgmNr 1348 A genome-wide survey of recurrent repeat expansions in cancer genomes

PrgmNr 1351 Brain-specific succinyl-CoA synthetase deficiency leads to increased protein succinylation, abnormal neuronal metabolism, and altered chromatin accessibility

PrgmNr 1352 Aberrant posttranslational modifications contribute to MMA pathophysiology and identify new targets for therapy

PrgmNr 1353 The gnomAD database of mitochondrial DNA variation across 56,434 individuals

PrgmNr 1354 Clinician lead analysis of WGS data identifies missed diagnoses in suspected primary mitochondrial disease

PrgmNr 1355 Regulating PCCA gene expression by splice switching antisense oligonucleotide-mediated modulation of pseudoexon splicing to rescue propionyl CoA carboxylase enzyme activity in propionic acidemia

PrgmNr 1356 Multi-omics and functional studies in two cellular models of Barth Syndrome reveal novel mitochondrial defects and suggest new therapeutic approaches

PrgmNr 1359 Systematic functional interrogation of the Lynch Syndrome gene MLH1 by deep mutational

PrgmNr 1360 Comprehensive cell-type classification of tumor and normal cells from single cell RNA sequencing in pan cancer settings

PrgmNr 1361 Deep transcriptomic profiling with quantitative measures identifies high-risk multiple myeloma

PrgmNr 1362 Integrated somatic and germline assessment of PROACTIVE genomic data

PrgmNr 1363 Germline HOXB13 variant contributes to risk of prostate cancer in men of African Ancestry

PrgmNr 1364 Identification of mitochondrial DNA variants associated with risk of neuroblastoma

PrgmNr 1367 The genomic landscape of mosaic Y chromosome loss in the VA Million Veteran Program

PrgmNr 1368 Genome-wide gene expression analyses reveals ancestry-specific genetic architecture in Latino and African American children

PrgmNr 1369 Single-cell RNA-sequencing reveals pervasive but highly cell type-specific genetic ancestry effects on the response to viral infection

PrgmNr 1370 Chromatin accessibility differences between African and European Alzheimer Disease brains

PrgmNr 1371 Protein prediction for trait mapping in diverse populations

PrgmNr 1372 Multi-ancestry GWAS for venous thromboembolism identifies novel loci followed by experimental validation

PrgmNr 1375 Quantifying the developmental trajectory of autism associated brain overgrowth using 3D cellular resolution imaging

PrgmNr 1376 Genomic architecture of autism spectrum disorder from comprehensive whole-genome sequence annotation

PrgmNr 1377 Genotype-phenotype analyses in 10,000 individuals with psychosis highlights complex role of ultra-rare coding variants and common polygenic risk in presentation and etiology

PrgmNr 1378 Genome-wide association study of over 40,000 cases of suicide attempts yields eight genome-wide significant loci

PrgmNr 1379 Chromatin accessibility mapping reveals compound genetic effects at a schizophrenia GWAS risk locus impairing neurodevelopment and synaptic function in human neurons

PrgmNr 1380 Lineage-specific analysis of epigenome and transcriptome changes in postmortem brains from Schizophrenia and Bipolar Disorder

Poster Talk

296 - 379

PrgmNr 1407 Cancer Risk C, a functional genomics test, is a sensitive, specific, and rapid diagnostic for Lynch syndrome

PrgmNr 1408 Does it make sense to report single pathogenic alleles in cancer susceptibility genes for childhood cancer patients? Results from the BASIC3 and Texas KidsCanSeq studies

PrgmNr 1409 Breast cancer polygenic risk scores and rare variants in Latinas

PrgmNr 1410 Germline genetically predicted mammographic density is associated with breast tumor transcriptomic and proteomic features, CD8+ T cell infiltration, somatic mutational signatures, and tumor mutational burden in The Cancer Genome Atlas

PrgmNr 1411 Saturation Genome Editing of PALB2 Reveals Functionally Abnormal Missense Variants

PrgmNr 1412 Identify non-mutational TP53 loss of function in human cancers

PrgmNr 1413 Multiple germline events lead to cancer development in patients with Li-Fraumeni syndrome

PrgmNr 1414 Oncogenic mutations in the normal human brain

PrgmNr 1415 A human importin-ß-related disorder: Syndromic thoracic aortic aneurysm caused by bi-allelic loss-of-function variants in IPO8

PrgmNr 1416 Genotype-phenotype correlations in PIK3CA-related overgrowth spectrum (PROS) and overlapping phenotypes: a systematic review of 1007 patients with PIK3CA pathogenic variants

PrgmNr 1417 Exome CNV calling and analysis in a large cohort of families with undiagnosed rare genetic disease

PrgmNr 1418 Brazil enters the genome sequencing era: The Rare Genomes Project - interim results

PrgmNr 1419 Low-pass genome sequencing-based detection of absence of heterozygosity: validation in clinical cytogenetics

PrgmNr 1420 Advanced Diagnostics and Genotype-Phenotype Resolution using Functional Genomics in >500 Neuromuscular and Neurological Disorder Patients

PrgmNr 1421 Functional characterization of haplotype surrounding TOMM40-523' repeat to assess differential risk effects on European ancestry APOEε3 haplotypes

PrgmNr 1422 Telomere length analysis in amyotrophic lateral sclerosis using large-scale whole genome sequence data

PrgmNr 1423 Shared genetic aetiology of type 2 diabetes and knee osteoarthritis

PrgmNr 1424 Genetic control of mRNA splicing is affected by purifying selection and may be linked to incomplete penetrance

PrgmNr 1425 Rare variants affecting multi-omic measurements of gene regulation implicated in EKG traits

PrgmNr 1426 Systematic evaluation and Improvements for trans-eQTL Detection Methods Allows Identification of Novel trans-eQTLs in the GTEx data

PrgmNr 1427 Brain eQTL of East Asian, African American, and European Descent Explains Schizophrenia GWAS in Diverse Populations

PrgmNr 1428 Genome-wide inter-chromosomal epistatic associations identified across complex diseases in the ~300,000 participants from eMERGE and UK Biobank

PrgmNr 1429 Global Biobank Meta-Analysis Initiative: A genome-wide association meta-analysis identifies novel primary open-angle glaucoma loci and shared biology with vascular mechanisms

PrgmNr 1430 Systematic comparison of family history and polygenic risk across 27 common diseases

PrgmNr 1431 Systematic comparison of performance of over 50 PRS across ancestries for 14 diseases

PrgmNr 1432 Power-improved meta-QTL analysis reveals the complex regulation of molecular QTL associated with brain diseases

PrgmNr 1433 Single-cell RNA sequencing of lung identifies disease-associated cell type-specific eQTLs

PrgmNr 1434 The Musculoskeletal 3D Epigenome Atlas

PrgmNr 1435 Enhancer enhancer interaction networks involved in long-distance genome regulation link multiple non-coding variants to function

PrgmNr 1436 Escape from X-inactivation in twins exhibits intra- and inter-individual variability across tissues

PrgmNr 1437 ChromBPNet: Deep learning models of base-resolution chromatin profiles reveal cis-regulatory syntax and regulatory variation

PrgmNr 1438 NCBI ALFA Release 2for900 Million Variantsand Allele Frequencyfrom 200KdbGaPSubjects

PrgmNr 1439 Phenome-wide association analysis of rare and common variation in 455,000 UK Biobank

PrgmNr 1440 Integrated multi-omics analysis of thyroid cancer reveals key molecular pathways involved in tumor formation and metastasis

PrgmNr 1441 Whole Blood RNA Sequencing in a Cohort of Undiagnosed Pediatric Patients

PrgmNr 1442 Long read sequencing reveals extensive structural variation in diverse mouse genomes

PrgmNr 1443 MERSCOPETM reveals the transcriptional organization of the mouse brain

PrgmNr 1444 Identifying the genetic contributors of efficacy and adverse metabolic effects of thiazide diuretics in African Americans from the Genetics of Hypertension Associated Treatments (GenHAT) study

PrgmNr 1445 Paradigm Shift from Disease PRS to PGx PRS for Drug Response Prediction using PRS-PGx Methods

PrgmNr 1446 Utility of polygenic risk scores for colorectal cancer risk assessment across diverse populations

PrgmNr 1447 Dissecting 90 percent of lipoprotein(a) heritability in 487,571 UK Biobank participants

PrgmNr 1448 Polygenic score from a large GWAS predicts cases of heart failure with reduced ejection fraction (HFrEF) but not preserved ejection fraction (HFpEF)

PrgmNr 1449 Rare variant carrier status on risk of coronary artery disease among individuals with high polygenic risk

PrgmNr 1450 Expanded Genetic Clustering of Type 2 Diabetes Loci Using a High-throughput Pipeline Reveals Mechanistic Pathways of Metabolic Diseases

PrgmNr 1451 Mirror effects of OPRD1 variants on diabetes and obesity

PrgmNr 1452 The effect of obesity-related traits on COVID-19 severe respiratory symptoms and hospitalization and its mediation by socioeconomic status: a multivariable Mendelian randomization study

PrgmNr 1453 Comparative genetic architecture of the inflammatory bowel disease across East Asian and European populations

PrgmNr 1454 High-resolution genomic architecture of COVID-19 severe disease using multi-ethnic whole genome sequencing data

PrgmNr 1455 A de novo paradigm for male infertility

PrgmNr 1456 Genomic and Transcriptomic-wide Analysis Identifies Novel Genetic Risk Loci and Prioritization of Therapies for Myasthenia Gravis

PrgmNr 1457 Polygenic risk scores as a marker for lifetime epilepsy risk

PrgmNr 1458 Trans-ancestry GWAS meta-analysis of alcohol and tobacco addiction in 3.4 million individuals

PrgmNr 1459 A Genealogical Estimate of Genetic Relationships to Improve Detection of Population Structure Over Time

PrgmNr 1460 DNA methylation patterns underlying lifespan differences in mammals

PrgmNr 1461 Guaranteeing unbiasedness in selection tests based on polygenic scores

PrgmNr 1462 A cross-disorder and evidence-based tiered ranking of candidate genes for neuropsychiatric disorders

PrgmNr 1463 Dysregulation of Canonical and Alternative Replication Protein A Complexes in Huntington Disease and Spinocerebellar ataxia Type 1 brains is associated with CAG instability and phenotype

PrgmNr 1464 Genotypic spectrum and its clinical implication in disorders with epilepsy in Indian population: A preliminary experience

PrgmNr 1465 Identification of NDEL1 as a novel gene for lissencephaly

PrgmNr 1466 EFEMP1 rare variants cause juvenile-onset open angle glaucoma in families from the Philippines

PrgmNr 1469 Genetics Adviser: The development, usability and acceptance testing of a patient-centered digital health application to support clinical genomic testing

PrgmNr 1470 The cost of good health: Poverty association with differential gene expression

PrgmNr 1471 The impact of clinically relevant CNVs in the general population - the health consequences and personalized management of undiagnosed adult CNV carriers in the Estonian biobank

PrgmNr 1472 Uptake of cancer risk management strategies among women who undergo cascade genetic testing for breast cancer susceptibility: How do they compare to non-cascade testers?

PrgmNr 1473 Haplotyping SNPs for allele-specific gene editing of the mutant huntingtin allele using long-read sequencing

PrgmNr 1474 A mouse model recapitulating Bruck Syndrome provides insight into the role of Plod2 in bone

PrgmNr 1476 Mutations in DNA ligase III cause mitochondrial neurogastrointestinal encephalomyopathy

PrgmNr 1477 Significant burden of de novo damaging variants in novel genes in patients with congenital kidney malformations

PrgmNr 1478 Genome-wide genetic control of fetal placental genomics shows multiple associations with health and disease across the life course, informing the Developmental Origins of Health and Disease

PrgmNr 1479 Residual risk for clinically significant copy number variants in pregnancies with normal NIPS

PrgmNr 1480 A 100,000 Genome Project haplotype reference panel of 156,390 haplotypes and the improved imputation of UK Biobank

PrgmNr 1481 A powerful test of ancestral heterogeneity in the effects of gene expression on complex traits

PrgmNr 1482 Developing Trans-ethnic Polygenic Risk Scores Using Empirical Bayes and Super Learning Algorithm

PrgmNr 1483 Explainable and extendable machine learning models for identifying prognostic radiogenomic biomarkers from breast cancer multimodal imaging and genomic data

PrgmNr 1484 Identity-by-descent mapping in biobank-scale datasets

PrgmNr 1485 Incorporating family disease history and controlling case-control imbalance for population based genetic association studies

PrgmNr 1486 SUMMIT: An integrative approach for better transcriptomic data imputation improves causal gene identification

PrgmNr 1487 Allele-specific expression of SNPs involved in the immune response and gene transcription is associated with BMI

PrgmNr 1488 Genetic determinants of prostate-specific antigen levels improve cancer screening utility

PrgmNr 1489 Joint intron splicing-based transcriptome-wide association study identifies new candidate susceptibility genes for breast cancer

Poster Presentations

380 - 2225

PrgmNr 2001 ANGPT1 TEK signaling pathway and primary congenital glaucoma A likely impact in a female newborn glaucoma associated to a dup8q22

PrgmNr 2002 Chromosomal abnormality -17/17P- in acute B-lymphoblastic leukemia (B-ALL) implies loss of heterozygosity of TP53 gene

PrgmNr 2003 Clinical features and genetic analysis of pediatric pilomixoid astrocytomas

PrgmNr 2004 CYP polymorphisms and heart failure

PrgmNr 2005 Geographic differences of the distribution of the mutational spectrum of BRCA1 and BRCA2 genes in Algerian population

PrgmNr 2006 Germline mitochondrial DNA non-coding region variants among sporadic breast cancer patients of two Sri Lankan ethnic groups

PrgmNr 2007 Identification of RP gene and ERα gene polymorphism as genetic markers for the improvement of breast cancer management in cameroon

PrgmNr 2008 Low Prevalence of Large Genomic Deletions and Duplications Observed in the Sequential Use of MLPA after Negative NGS Result in Thai Breast Cancer Cohort

PrgmNr 2009 Renal cancer progression & contribution of chromatin remodeling complexes

PrgmNr 2010 Small nucleolar RNAs as potential predictive/prognostic markers for breast cancer

PrgmNr 2011 Studying the phenotypic variability of ERCC6L2 deficiency in a genetic island

PrgmNr 2012 The Immune-genetics landscape of brain metastases compared to their primary Tumors in a Saudi population

PrgmNr 2013 Timer CAR-T (TCAR-T): A paradigm shift in CAR-T cell therapy

PrgmNr 2014 A prospective Study Using Multigene Next Generation Sequencing panel to Evaluate the Prevalence of Familial Cancer in a Highly Consanguineous Population

PrgmNr 2015 Genetic factors of differentiated thyroid cancer in French Polynesian population after exposure to nuclear tests, A suggested role of 3 loci

PrgmNr 2016 Hereditary Bilateral Breast Cancer: role of the Multi-Gene Panel Testing for detection of Pathogenic Variants

PrgmNr 2017 Statistical Efficiency of Meta-analyses in Testing Gene-environment Interactions with Rare Variants

PrgmNr 2018 DNA repair genes in polyposis susceptibility

PrgmNr 2019 Bioinformatics analysis of gene expression profile in haematologic malignancies

PrgmNr 2020 Cancer informatics survey of different grades and subtypes of glioma

PrgmNr 2021 Chromatin domains affect the clustered breakpoint formation in complex chromosomal rearrangement

PrgmNr 2022 Deep generative neural network for accurate drug response imputation

PrgmNr 2023 Distinctive miRNA expression profiling of pediatric solid tumors in relation to doxorubicin treatment

PrgmNr 2024 Exploration of hypomorphic variants from VUS in multigene panel testing for Japanese breast cancer patients

PrgmNr 2025 Germline landscape of BRCAs by 7-site collaborations as a BRCA consortium in Turkey

PrgmNr 2026 HIVID2: an accurate tool to detect virus integrations in the host genome

PrgmNr 2027 Mutation profile of circulating cell-free DNA reflects the evolution of metastatic breast cancer

PrgmNr 2028 Reversal of Anthracycline Resistance in Colon Cancer cell lines using natural compound Curcumin

PrgmNr 2029 Single cell transcriptomics of Pituitary Neuroendocrine Tumors (PitNETs)

PrgmNr 2030 Spatial transcriptomics reveals unique molecular features of fluorescence sorted 5 minolevulinic acid positive infiltrative tumor cells associated with recurrence and poor survival in glioblastoma

PrgmNr 2032 LPA, KIV2, serum Lp(a), and the risk for Cancer and Cardiovascular disease related pathologies: A retrospective autopsy study

PrgmNr 2033 Trans-ancestry GWASs for ECG markers of ventricular depolarization and repolarization in 250,730 individuals identifies shared and distinct mechanisms

PrgmNr 2034 A Novel Method of Identifying Inherited High-risk Population of Atherosclerotic Cardiovascular Disease Using Polygenic Risk Scores of Metabolic Diseases

PrgmNr 2035 A translational bioinformatics investigation of the human gut microbiome and hypertensive diseases

PrgmNr 2036 Genome-wide association study of lipids in Greenlandic Inuit show large-effect size and independent associations and reveals a unique genetic architecture

PrgmNr 2037 Integration of biomarker polygenic risk score improves the prediction of coronary artery disease in UK Biobank and FinnGen

PrgmNr 2038 Multivariate genetic analysis of human plasma glycerophospholipids, glycerolipids, sphingolipids and sterols identifies novel associations near LPGAT1, GRIP1, SGPL1 and ERMP1

PrgmNr 2039 Rare Variant Analysis for Coronary Heart Disease Cases and Controls using Whole Genome Sequencing in a Middle Eastern Population Identifies Potential Novel Genes

PrgmNr 2040 Sex-stratified whole genome sequencing association study for coronary heart disease in a middle eastern population suggests differential loci

PrgmNr 2041 The penetrance of rare cardiomyopathy-associated DNA variants: a cross-sectional approach to estimate penetrance using large case series

PrgmNr 2042 Transferability of genetic loci and polygenic scores for cardiometabolic traits in ~22,000 British Pakistani and Bangladeshi individuals from a real-world healthcare cohort

PrgmNr 2043 Whole Genome Sequencing Association Study for Coronary Heart Disease in a Middle Eastern Cohort Validates Polygenic Risk Scores and Replicates Known Loci

PrgmNr 2044 A whole genome sequencing study in a three-generation consanguineous family with morbid obesity, learning difficulty and failure of weight loss surgery, to identify susceptibility variants for obesity

PrgmNr 2045 Deep shotgun metagenomic sequencing and functional profiling of the gut microbiota identifies associations with kidney function

PrgmNr 2046 Estimation of variance explained by polygenic risk score of Crohn's disease and ulcerative colitis using Korean versus European GWAS

PrgmNr 2048 HLA haplotype determines treatment response to GAD-alum immunotherapy in Type 1 diabetes: An applied example of a genetic precision medicine approach to pivotal clinical trial design

PrgmNr 2049 Selection pressures affecting Type 2 Diabetes in a South Indian and Scottish population

PrgmNr 2050 Validation and risk score applicability of T2D related variants in populations of East Asians and Europeans

PrgmNr 2051 Whole genome analysis of 342 adult cases of obesity and related metabolic disorders in highly consanguineous population

PrgmNr 2052 HLA-A*11:01:01:01, HLA*C*12:02:02:01-HLA-B*52:01:02:02, age and sex are associated with severity of Japanese COVID-19 with respiratory failure

PrgmNr 2053 Applying the regional heritability mapping method to primary biliary cholangitis in the Japanese population

PrgmNr 2054 Common, low frequency, rare and ultra-rare variants contribute to COVID-19 severity

PrgmNr 2055 Comprehensive analysis of HLA association in Japanese with childhood-onset nephrotic syndrome

PrgmNr 2056 Effective way that determining mild or severe COVID-19 patient only using T-cell receptor(TCR) sequencing data

PrgmNr 2057 Fine-mapping of novel susceptibility loci associated with eosinophil granule proteins (ECP and EDN) reveals five biologically relevant genes for asthma

PrgmNr 2058 Genetic diversity and epidemiology of human rhinovirus among children with severe acute respiratory tract infection in Guangzhou of China

PrgmNr 2059 Genetic Similarity Assessment of Qatar and Italy Populations in the context of COVID-19

PrgmNr 2060 High-resolution genomic architecture of COVID-19 severe disease using multi-ethnic whole genome sequencing data

PrgmNr 2061 Identification of causal genetic variants in invasive pneumococcal diseases by exome analysis in children

PrgmNr 2062 Identifying novel causative mutations forDOCK8 immunodeficiency syndromeusing whole exome sequencing

PrgmNr 2063 PADI4 and PADI2 enhance collagen-initiated inflammatory responses

PrgmNr 2064 rs1944919 on human chromosome 11q23.1 and its effector genes COLCA1 and COLCA2 confer susceptibility to primary biliary cholangitis

PrgmNr 2065 The impact of SARS-CoV-2 multiple spike protein variants on COVID-19 outcomes

PrgmNr 2066 Whole Exome Sequencing for Covid19 host in mild versus severe cases: A Saudi genome study

PrgmNr 2067 Advancing our understanding of genetic risk factors and potential personalized strategies in pelvic organ prolapse: currently largest GWAS reveals 19 novel associations

PrgmNr 2068 Explore the dynamic spatial and temporal regulation mechanism during embryonic sex development

PrgmNr 2069 Genetic characterization of the timing of human parturition unveils differential roles of maternal and fetal genomes on birth weight

PrgmNr 2070 Genome-wide polygenic risk scores for hypertensive disease during pregnancy identify women at risk for long-term cardiovascular disease

PrgmNr 2071 Large-scale GWAS meta-analysis unravels the genetic determinants of cervical biology and pathology

PrgmNr 2072 Meta-analysis of genome-wide association studies identified new loci associated with spontaneous preterm birth and gestational duration

PrgmNr 2073 Novel bi-allelic Mutations in DNAH10 cause multiple morphological abnormalities of the flagella

PrgmNr 2074 Disentangling migraine risk loci by fine-mapping and colocalization

PrgmNr 2076 Long-read sequencing analysis of Alzheimer risk geneABCA7 identifies risk-increasing novel alternative splicing events

PrgmNr 2077 Polygenic risk scores as a marker for lifetime epilepsy risk

PrgmNr 2078 Protein interaction network reveals enriched genetic variation across signaling networks in frontotemporal dementia

PrgmNr 2079 Trans-ethnic fine-mapping in the major histocompatibility complex region on Parkinson's disease risk

PrgmNr 2080 Using non-Autologous Wharton's Jelly Mesenchymal Stem Cells and Chitosan/Poly ethylene oxide as a Synthetic Scaffold inhibit SNI-Induced Apoptosis in Rat

PrgmNr 2082 Assessing association between polygenic risk scores for mood disorders and substance involvement in the Taiwan Biobank

PrgmNr 2083 Brain eQTL of East Asian, African American, and European Descent Explains Schizophrenia GWAS in Diverse Populations

PrgmNr 2084 Contribution of Glucoside Xylosyltransferase 1 gene variants in Schizophrenia

PrgmNr 2086 Is chronotype a risk factor for neuropsychiatric disorders? A two-sample, multivariable Mendelian randomisation study

PrgmNr 2087 Major depressive disorder and current psychological symptoms modify the polygenic predisposition to body mass index on obesity-related traits

PrgmNr 2088 A phenome-wide association study of Y-chromosomal haplogroups and 28 disease endpoints in 37,518 Finnish men

PrgmNr 2089 A selection pressure landscape for 870 human polygenic traits

PrgmNr 2090 Analysis of genetic variants of SARS-CoV-2 host factors ACE2, NRP1, TMPRSS2, and FURIN in Japanese population from IRUD whole-exome sequencing data

PrgmNr 2091 Analysis of the genetic components behind the chronic back pain

PrgmNr 2092 Association of confounding factors with Adolescent Idiopathic Scoliosis in the population of Jammu andKashmir, India

PrgmNr 2093 Can sex differences in GWAS variant effects be found with current sample sizes?

PrgmNr 2094 Challenges with X chromosome analyses and reporting in Genome-Wide Association Studies

PrgmNr 2095 Comprehensive genome-wide association study of different forms of hernia identifies more than 80 associated loci

PrgmNr 2096 Deep integrative models for large-scale human genomics

PrgmNr 2097 Genome-wide association study of chronotype in Japanese direct-to-consumer genetic testing data

PrgmNr 2098 Genome-wide association study of varicose veins identifies a protective low-frequency missense variant in GJD3 enriched in Finnish population and highlights connexins as potential therapeutic targets

PrgmNr 2099 Genome-wide phenome-wide association study for important diseases & quantitative traits based on the Taiwan Biobank data
PrgmNr 2100 Genotype-by-environment interactions for chronic back pain
PrgmNr 2101 Hierarchical topic modelling enables genetic analysis of electronic healthcare record data
PrgmNr 2102 Machine-learning SNP-based prediction for Primary Biliary Cholangitis: A proof-of-concept study
PrgmNr 2103 Phenome-wide association study of deleterious variant (shet) burden using exome sequencing of 394,694 UK Biobank participants
PrgmNr 2105 Quantifying concordant genetic effects of de novo mutations on multiple disorders
PrgmNr 2106 Sources of Inequality at Birth: The Interplay Between Genes and Parental Socioeconomic Status
PrgmNr 2107 Systematic comparison of family history and polygenic risk across 27 common diseases
PrgmNr 2108 Systematic comparison of performance of over 50 PRS across ancestries for 14 diseases
PrgmNr 2109 Web-based, participant-driven research platform discovered novel genetic associations for stress-related complaints and food-related behaviors in Japanese population
PrgmNr 2110 Epigenome-wide association study of DNA methylation in CD4- and CD8-positive T cells in narcolepsy
PrgmNr 2111 Functional characterisation of GJB2 cis-regulatory elements and WGS of heterozygous patients with NSHL
PrgmNr 2112 Genome-wide Analysis of DNA methylation in 106 Schizophrenia family trios in Han Chinese
PrgmNr 2113 Genome-wide DNA methylation profiling of leukocytes identifies CpG methylation signatures of treatment-resistant schizophrenia
PrgmNr 2114 High genetic load of the GREB1L gene and auditory rehabilitation in the severe form of cochlear malformation
PrgmNr 2115 Mapping the cis-regulatory landscape of ABCA4 in adult human retina
PrgmNr 2116 Methylation quantitative trait loci of type 2 diabetes in a middle eastern population
PrgmNr 2117 Mitochondrial-nuclear gene expression co-regulation is compromised in immune system cells of COVID-19 patients: rewiring towards glycolysis
PrgmNr 2118 UMI-4C chromatin interaction profiling, in vitro and in vivo enhancer assays to dissect the cis-regulatory mechanisms underlying North Carolina macular dystrophy, a retinal enhanceropathy
PrgmNr 2119 ZNF143, an important factor for CTCF-bound chromatin interactions
PrgmNr 2120 Epigenome-wide association study of the human IgG N-glycome composition
PrgmNr 2121 eQTL Catalogue: A compendium of uniformly processed human gene expression and splicing QTLs
PrgmNr 2122 Evaluation of the impact of genetic diversity on PIWI-interacting RNA analysis
PrgmNr 2123 Loss of PTEN disrupts 5-hydroxymethylcytosine landscape along mESC neuronal differentiation
PrgmNr 2124 Nuclease deficiencies alter plasma cell-free DNA methylation profiles
PrgmNr 2125 Regulon active landscape reveals cell development and function state changes of human primary osteoblasts in vivo
PrgmNr 2126 TDG contributes to male subfertility through dysregulation of male germline stem cell development
PrgmNr 2127 Whole-genome and small RNA sequencing-based microRNA-eQTL mapping in Japanese elucidates variant-microRNA-disease connections
PrgmNr 2128 Identification of hubb genes in retinoblastoma using network analysis
PrgmNr 2129 Detecting gene-gene interactions from GWAS using diffusion kernel principal components
PrgmNr 2130 Epigenetic and RNA profiling of cancer using multi-omics data on the Cancer Genomics Cloud powered by Seven Bridges
PrgmNr 2131 Implementation and evaluation of available methods for epimutation analysis

PrgmNr 2132 Pathway analysis enhances characterization of cell types and sample groups in single-cell RNA sequencing

PrgmNr 2133 Spatial changes in the human genome indicated by the consensus-based structural variants in selected human families

PrgmNr 2134 CNV characteristics in Chernobyl power plant catastrophe clean-up workers from Lithuania suggest unique genetic variation structure

PrgmNr 2135 Differentiation in Olfactory Receptor Genes in Worldwide Populations

PrgmNr 2136 Evaluating the accuracy of genotype imputation in the MHC region in selected African populations

PrgmNr 2137 Genetic diversity in host immune genes involved in responses against intestinal parasitic infections in the Orang Asli community in Malaysia

PrgmNr 2138 Genomic and ancestral variation underscores the severity of COVID-19 disease presentation

PrgmNr 2139 Gut microbiome analyses of ancient individuals, so called “Jomon”, lived in Japanese archipelago

PrgmNr 2140 Hematological, epigenetic and genetic adaptations in Tibetans

PrgmNr 2141 Historical Demography of Lithuania and Relationship to other Populations

PrgmNr 2142 Incidence of patients with methylenetetrahydrofolate reductase (MTHFR) gene mutations in a Tunisian population cohort

PrgmNr 2143 Population-specific adaptation to malaria infection in endemic regions of Asia

PrgmNr 2144 Positive selection to altitude and phenotypic associations in Papua New Guinean highlanders

PrgmNr 2145 TogoVar; providing integrated genomic information of Japanese population

PrgmNr 2146 Variations in genes associated with recurrent miscarriages in the HGDP CEPH dataset

PrgmNr 2147 Classifying uncertainty and impact of uncertainty on decision-making: How do medical students handle uncertainty in prenatal exome sequencing?

PrgmNr 2148 Population biobank participants’ response to a range of genomic results reported

PrgmNr 2149 Psychological burden of preimplantation genetic testing (PGT) on couples with multiple monogenic disorders and the role of genetic counselling in Saudi Arabia

PrgmNr 2150 Role of whole genome sequencing in detecting compound heterozygotes for single nucleotide variant and structural variant: Two illustrative patients

PrgmNr 2151 Stigma-power around Fragile X Syndrome: a tale of a royal family and a community in a rural village in Cameroon

PrgmNr 2152 The impact of clinically relevant CNVs in the general population - the health consequences and personalized management of undiagnosed adult CNV carriers in the Estonian biobank

PrgmNr 2153 Use of MOOCs for inclusive, flexible professional development and wider public engagement in human genomics and genetics

PrgmNr 2154 Novel CRISPR system with modified cas12 and nanoliposome based delivery can destroy SARS covid2 virus in human lung tissue

PrgmNr 2155 Targeted Therapies for Hereditary Peripheral Neuropathies: Systematic Review and Steps Towards a 'treatabolome'

PrgmNr 2156 A global omics data sharing and analytics marketplace: Case study of a rapid data COVID-19 pandemic response platform

PrgmNr 2157 An atlas of associations between polygenic risk scores from across the human phenome and circulating metabolic biomarkers

PrgmNr 2158 Human genetics is great: A genomic structural variant map of 945 Han Chinese individuals using long-read sequencing data

PrgmNr 2159 Measuring sensitivity of semi-automated Clingen-ACMG classification framework for CNVs through comparison with manual classification of 1437 clinical cases

PrgmNr 2160 SpliceAI algorithm identified potentially pathogenic variants in patients with hereditary hearing impairment

PrgmNr 2161 Towards a more usable database: MGeND 2021 update

PrgmNr 2162 Variations in Nomenclature of Clinical Variants between Annotation Tools

PrgmNr 2163 Two new patients with focal dermal hypoplasia: A novel PORCN variant and insights on the differential diagnostic considerations

PrgmNr 2164 A human importin-ß-related disorder: Syndromic thoracic aortic aneurysm caused by bi-allelic loss-of-function variants in IPO8

PrgmNr 2165 A novel NOTCH3 terminal exon variant causes the rare lateral meningocele syndrome in an Asian child

PrgmNr 2166 A novel variant in the WNT10A gene in a consanguineous Malian family with Odonto-Onycho-Dermal Dysplasia syndrome: First sub-Saharan African case

PrgmNr 2167 A Palestinian patient with Wiedemann-Rautenstrauch like syndrome: expanding the phenotypic spectrum of PYCR1 mutations

PrgmNr 2168 Acromesomelic skeletal dysplasia with severe short stature due to a biallelic KIF24 variant

PrgmNr 2169 Clinically actionable secondary findings in 390 whole genome sequence data from sub-Saharan African families with nonsyndromic orofacial clefts

PrgmNr 2170 Congenital bilateral anophthalmia: A case report

PrgmNr 2171 C-terminal and N-terminal truncating mutations of the MN1 gene lead to distinct developmental phenotypes

PrgmNr 2172 Delineation of the phenotype and genotype in ten individuals with de novo variants in ZBTB18

PrgmNr 2173 Expanding the phenotype of cerebello-facio-dental syndrome in female with a novel pathogenic variant

PrgmNr 2174 Long-read whole genome sequencing identified a partial MBD5 deletion in an exome-negative patient with neurodevelopmental disorder

PrgmNr 2175 Loss-of-Function mutations of USP9X as the differential diagnosis for CHARGE syndrome

PrgmNr 2176 Objective evaluation of dysmorphism by automated analysis of facial photographs

PrgmNr 2177 Say-Barber-Biesecker-Young-Simpson syndrome (SBBYSS) and the importance of Whole Exome Sequencing in finding the accurate diagnostic. A case report

PrgmNr 2178 The first East Asian adult case with rare NEPROrelated skeletal dysplasia caused by a novel missense homozygous variant in NEPRO

PrgmNr 2179 The genotypic and phenotypic spectrum of pycnodysostosis in Saudi Arabia: Novel variants and clinical findings

PrgmNr 2182 A novel variant in the GNE gene in a Malian with distal muscle weakness

PrgmNr 2183 A rare double diagnosis identified via exome sequencing in a patient with complex cerebellar ataxia

PrgmNr 2184 An exonic LINE-1 insertion and a novel missense variant in CC2D2A identified in siblings with Joubert syndrome using Long-Read Sequencing

PrgmNr 2186 Bi-allelic variants in neuronal cell adhesion molecule (NRCAM) lead to a novel neurodevelopmental disorder characterized by developmental delay, hypotonia, peripheral neuropathy or spasticity

PrgmNr 2187 Comprehensive genetic investigation of penta-nucleotide tandem repeats at RFC1 locus in Indian ataxia and ALS cohort

PrgmNr 2188 Delineating the spectrum of disorders with CNS white matter abnormalities in Indian population

PrgmNr 2189 Delineation of the phenotypic and genotypic spectrum of disorders with deficient myelination of central nervous system in 26 Indian families

PrgmNr 2190 Expanding the genotypic spectrum of Allan-Herndon-Dudley syndrome: Report of two novel mutations in the SLC16A2 gene

PrgmNr 2191 Genotypic spectrum and its clinical implication in disorders with epilepsy in Indian population: A preliminary experience

PrgmNr 2192 Identification of NDEL1 as a novel gene for lissencephaly

PrgmNr 2193 LRFN1- a novel candidate gene as a potential cause of autism spectrum disorder

PrgmNr 2194 Metabotropic glutamate receptor subtype 2 (GRM2) is a novel disease gene for developmental and epileptic encephalopathy

PrgmNr 2195 Mono- and bi-allelic variation of ABCA2 in individuals with different neurodevelopmental disorders

PrgmNr 2196 Multiple methodologies reveal novel picture of the genetic architecture of intellectual disability in Northern Finland

PrgmNr 2197 Mutational Landscape of a Bangladeshi Cohort of Neurodevelopmental Disorders

PrgmNr 2198 Novel PCDH19 frameshift variant in a girl with recurrent seizures and phenotype overlapping with SCN1A gene dysfunction: a case report and literature review

PrgmNr 2199 Preliminary studies on apparent Mendelian psychotic disorders in consanguineous families

PrgmNr 2200 Similar genotypes leading to different phenotype: expansion of the phenotype of UDP-Glucose-6-Dehydrogenase recessive neurodevelopmental disorder with or without epileptic encephalopathy, 2 new descriptions

PrgmNr 2201 The Genetic Architecture of Neurological Diseases in Khyber Pakhtunkhwa-Pakistan

PrgmNr 2202 A large deletion in EVER1 gene: Identification and validation through amplicon-based Next Generation Sequencing

PrgmNr 2203 A novel deletion causes SLC25A42-associated mitochondrial encephalomyopathy in a Saudi patient: Report of additional founder cases and functional characterization study

PrgmNr 2204 An MRM2 splicing variant in a complex dystonic syndrome: a second report and incomplete penetrance

PrgmNr 2205 Associate professor of clinical genetics

PrgmNr 2206 Bi-allelic missense variant p.Thr368Ala in DLST is associated with 2-oxoglutarate dehydrogenase complex deficiency related neurometabolic disorder

PrgmNr 2207 Drug repurposing based on multi-omics data for osteoporosis

PrgmNr 2208 Exome sequencing in 13 families with monogenic forms of rickets

PrgmNr 2209 Further delineation of RPL13-related skeletal dysplasia

PrgmNr 2210 Genetic linkage analysis identifies nuclear regions that modify the phenotype of m.3243A>G-related mitochondrial disease

PrgmNr 2211 High prevalence of DMD gene mutations in Duchenne muscular dystrophy in South Indian population

PrgmNr 2212 Identification of novel candidate genes and expansion of the phenotypic spectrum in a large skeletal dysplasias cohort

PrgmNr 2213 Long-term prognosis and genetic background of cardiomyopathy in 223 mitochondrial disease children

PrgmNr 2214 Lysosomal storage disorders at a tertiary care centre in a developing country

PrgmNr 2215 Mechanism of WRN loss for causing short stature in Werner syndrome

PrgmNr 2216 Molecular diagnostic challenges related to mutations in the low-complexity domain of heterogenous nuclear ribonuclearprotein A1 (HNRNPA1)

PrgmNr 2217 Mutations in DNA ligase III cause mitochondrial neurogastrointestinal encephalomyopathy

PrgmNr 2218 NTBC efficacy in alkaptonuria

PrgmNr 2219 Outcome of Clinical Genetic Testing in Patients with Features Suggestive for Ehlers Danlos Syndrome

PrgmNr 2220 Phenotypic Spectrumof DihydrolipoamideDehydrogenase Deficiencyin Saudi Arabia

PrgmNr 2221 The p.L3P (p.Leu3Pro) variant in GLA is not associated with Fabry disease

PrgmNr 2222 Association analysis of rare variants identified risk genes with neonatal respiratory distress syndrome

PrgmNr 2223 Combined immunodeficiency and increased cellular sensitivity to radiation due to a novel DNA Ligase 1 mutation

PrgmNr 2224 Functional characterization of a Xenopus tropicalis knockout and a human cellular model of RCBTB1-associated inherited retinal disease shows involvement of RCBTB1 in the cellular response to oxidative stress

PrgmNr 2225 Genetic and clinical characteristics of inherited retinal dystrophies in Northern Finland

PrgmNr 2226 Genetic variants influencing penetrance to iron overload disorder hereditary haemochromatosis: evidence from UK Biobank

PrgmNr 2227 Identification of variants associated with moderate to severe hearing loss

PrgmNr 2228 Inborn errors of immunity: Novel insights from four unrelated Indian families

PrgmNr 2229 Progressive familial intrahepatic cholestasis type3: two novel pathogenic variants in a cohort of Egyptian children with cholestatic disorders of infancy

PrgmNr 2230 Targeted exome sequencing identifies novel variants associated with non-syndromic hearing loss in an Indian family

PrgmNr 2231 A Novel Variant in RAP1GDS1 gene in a child with developmental

PrgmNr 2232 Beta-Ketothiolase deficiency in a patient with coexistence of homozygous ACAT1 gene mutation and homologous constitutional translocation t(4;11)(q13;q23) and father with XYY syndrome

PrgmNr 2233 Biallelic VPS35L pathogenic variants cause 3C/Ritscher-Scinzel-like syndrome: Description of two novel cases confirming the pathogenicity and clinical diversity

PrgmNr 2234 First reanalysis highlighting the increasing role of multiple molecular diagnoses in the field of rare diseases

PrgmNr 2235 Helsmoortel-van der Aa syndrome in a 14-year-old girl demonstrates intellectual impairment, facial dysmorphism and morbid obesity

PrgmNr 2236 Importance of genetic diagnosis on the PICU: case series during COVID-19 pandemic

PrgmNr 2237 MLPA mediated molecular characterization of associated syndromes with orofacial cleft patients

PrgmNr 2238 Pulmonary valvular stenosis missed Rasopathies in two asymptomatic adults

PrgmNr 2239 The identification of microduplications in 17q21.31 locus that might disrupt the KANSL1 gene

PrgmNr 2240 Clinical whole genome sequencing as an efficient first-tier diagnostic test for patients withrare diseasesfrom a resource-limited setting in the Democratic Republic of Congo

PrgmNr 2241 Current status of genetic diagnosis laboratories and frequency of genetic variants associated with cystic fibrosis through a newborn-screening program in Turkey

PrgmNr 2242 Detection of pathogenic CNVs in isolated and non-isolated congenital heart defects (CHD) by MLPA and CMA (retrospective cohort study)

PrgmNr 2243 Genetic determinants of mosaic loss of the X chromosome in peripheral leukocytes of 395,036 women from 3 biobanks

PrgmNr 2244 Low-pass genome sequencing-based detection of absence of heterozygosity: validation in clinical cytogenetics

PrgmNr 2245 Multilocus disease-causing genomic variations for Mendelian disorders: role of systematic phenotyping and implications on genetic counselling

PrgmNr 2246 The high-risk phenotypes of genetic disease in a Neonatal Intensive Care Unit population from the China neonatal genomes project

PrgmNr 2247 A novel splice site variant causes SLC13A5-related developmental and epileptic encephalopathy in large consanguineous family

PrgmNr 2248 Apolipoprotein E4 and meningeal lymphatics in Alzheimer disease: a conceptual framework

PrgmNr 2249 ----ASL deficiency in ALDH1A1+- neurons in the substantia nigra metabolically promotes neurodegenerative phenotypes

PrgmNr 2250 Compound heterozygousATMvariants cause late onset cerebellar and extrapyramidal disease in the absence of telangiectasia in a consanguineous Pakistani family

PrgmNr 2251 Dysregulated expression levels of APH1B in peripheral blood are associated with brain atrophy and amyloid-β deposition in Alzheimer’s disease

PrgmNr 2252 IMMP2L gene and Gilles de la Tourette syndrome a new case & review of the literature

PrgmNr 2253 Lipid transporter TMEM24/C2CD2L plays critical roles in ensuring the survival of retinal rod cells

PrgmNr 2254 Novel variants in UBE3B lead marfonoid body habitus of blepharophimosis ptosis intellectual disability syndrome

PrgmNr 2255 Phenotypic variability of MEGF10 variants in a consanguineous population

PrgmNr 2256 Second report of SHMT2 related neurodevelopmental disorder with cardiomyopathy, spasticity, and brain abnormalities

PrgmNr 2257 Simultaneous screening for SMN copy numbers and 2+0 silent carrier genotypes for Spinal Muscular Atrophy

PrgmNr 2258 The suppressive role for the splicing regulator, PTBP1, in the production of neuron-specific Agrin isoform

PrgmNr 2259 Transcriptomic profiling of chromatin-related neurodevelopmental disorders in human induced pluripotent stem cell -based models

PrgmNr 2260 Ultra rare truncating mutations of GRIK family genes associated with schizophrenia disrupt the interaction with PSD95 protein

PrgmNr 2261 Whole-genome sequencing analysis identifies rare Alzheimer's disease risk variants among Chinese individuals

PrgmNr 2262 Microtubule-Associated Serine/Threonine Kinases family genes (MAST) involved in the pathogenesis of seborrheic dermatitis may be related to Parkinson disease

PrgmNr 2263 RNAseq characterization of the effect of titin truncating variants

PrgmNr 2264 Automated prediction of the clinical impact of copy number variants: The power of combining expert and machine learning approach

PrgmNr 2265 Coding DNA numbering errors occurring in multiple pipelines

PrgmNr 2266 Copy number variation analysis with targeted next generation sequencing in patients with inherited metabolic disorders

PrgmNr 2267 Evaluating RNA-seq gene expression profiles for disease prediction using machine learning

PrgmNr 2268 Human Genome Topology for Selected Trios from the 1000 Genomes Project

PrgmNr 2269 Mining risk regulatory variants of Tetralogy of Fallot using deep learning models

PrgmNr 2270 Reanalysis of population cohort WGS revealed considerable minor allele frequency differences

PrgmNr 2271 VulExMap: Detection of exons which are vulnerable to exonic splicing mutations

PrgmNr 2272 A metagenome-wide association study revealed disease-specific landscape of the gut microbiome of systemic lupus erythematosus in Japanese

PrgmNr 2273 A multi-disciplinary approach to solving undiagnosed patients - Unsolved Cases Unit Groningen

PrgmNr 2274 An atlas of genetic scores to predict multi-omic biomolecular traits in blood

PrgmNr 2275 Enzymatic DNA synthesis (EDS) enables rapid and broad-based access to synthetic oligos needed for the genetic analysis and functional characterization of the SARS-CoV-2 virus

PrgmNr 2276 Impaired type I interferon activity of subjective cognitive decline in preclinical Alzheimer's disease

PrgmNr 2277 Molecular Cartography: a multiplexed high resolution transcriptomics approach to spatially analyze rare events demonstrated in an infection model of SARS-CoV-2

PrgmNr 2278 Translational diagnostics program -TDP-, an innovative intramural approach for hospital-based functional genomics of undiagnosed and rare diseases

PrgmNr 2279 An automated workflow for high-throughput qPCR

PrgmNr 2280 An integrated pipeline for genome analysis and phenotype-based gene prioritization with GenDiseak

PrgmNr 2281 Implementation of heart disease-associated cellular heterogeneity in vivo with Gene-Disease Integrative Systems Transgenesis (GENISYST) and Molecular Cartography

PrgmNr 2282 scDetect: A rank-based ensemble learning algorithm for cell type identification of single-cell RNA sequencing in cancer

PrgmNr 2283 Allelic and genotypic frequencies of CYP2C19*2, CYP2C19*3, and CYP2C19*17 alleles in the healthy South Indian population and their phenotypic prediction

PrgmNr 2284 Assessment of cytochrome P450 polymorphism for personalized therapy in acute coronary syndrome

PrgmNr 2285 Association between HLA-DRB1*15:01 and DQA1*01:02 frequencies and Lumiracoxib-induced DILI in healthy Thai Population

PrgmNr 2286 Association of a single nucleotide polymorphism rs12496846 in the C3orf20 gene region, identified to be associated with postoperative analgesia after mandibular sagittal split ramus osteotomy, with postoperative analgesia after major abdominal surgery

PrgmNr 2287 Comprehensive analysis of the genetics of clozapine-induced adverse effects

PrgmNr 2288 CYP2D6 genetic variants and their metabolic efficacy - insight from Molecular Dynamics Simulations

PrgmNr 2289 Distribution of HLA-B*13:01 allele related with dapsone-induced severe cutaneous adverse reaction in Thai and Asian population

PrgmNr 2290 Distribution of HLA-B*58:01 allele associated with allopurinol-induced SCARs in healthy Thai population

PrgmNr 2291 Experience and expectations of pharmacogenetic tests in France

PrgmNr 2292 Integration of genetically regulated gene expression and pharmacological library provides therapeutic drug candidates

PrgmNr 2293 Pharmacogenomics Marker of HLA-A*33:01 Allele Frequency in Healthy Thais

PrgmNr 2294 Practical guidelines of genomics-driven drug discovery from Global Biobank Meta-analysis Initiative

PrgmNr 2295 The frequency of HLA-B*57:01 allele in healthy Thai population

PrgmNr 2296 Association between relative telomere length in umbilical cord tissue and PCB congeners in maternal blood and cord blood in a birth cohort in Chiba, Japan

PrgmNr 2297 Distinct roles of ADAM10 and ADAM17 for testicular germ cell apoptosis

PrgmNr 2298 Exploring the factors affecting classification and reporting uncertain CMA findings, using a "virtual fetus" model

PrgmNr 2299 Genome-wide association analysis implicates dysregulation of Wnt signaling pathway in biliary atresia

PrgmNr 2300 Genomic sequencing in critically ill newborn infants shows high diagnostic rate in neurometabolic phenotypes and positive impact on clinical management

PrgmNr 2302 Residual risk for clinically significant copy number variants in pregnancies with normal NIPS

PrgmNr 2303 The potential of RNAsequencing for fetal malformation syndromes

PrgmNr 2304 3DeepMHC : 3D structure based Deep learning model for MHC-peptide binding prediction

PrgmNr 2305 A 100,000 Genome Project haplotype reference panel of 156,390 haplotypes and the improved imputation of UK Biobank

PrgmNr 2306 A weighted selection probability to locate rare variants associated with highly correlated multiple phenotypes

PrgmNr 2307 Accurate imputation of human leukocyte antigens with CookHLA

PrgmNr 2308 An explainable machine learning method for polygenic scores using whole UK biobank genotype dataset

PrgmNr 2309 Association tests of gene-environment interactions for time-to-event traits in family studies

PrgmNr 2311 Correction for sample overlap, winner's curse and weak instruments bias in two-sample Mendelian Randomization

PrgmNr 2313 Genome-wide classification of epigenetic signal reveals regions of enriched heritability in complex immune traits

PrgmNr 2314 Integrating facial and genomic data for subtyping individuals in the presence of confounders via individual-specific network analysis

PrgmNr 2315 Mendelian Randomization with repeated measurement using Taiwan Biobank

PrgmNr 2316 Significance in individual-specific networks: assessment, methods and characteristics

PrgmNr 2317 SNPs tagging structural variations help to reveal widespread role of CNVs in complex traits and gene expression

PrgmNr 2318 Statistical selection method to identify pleiotropic variants associated with both quantitative and qualitative traits

PrgmNr 2319 The Alpha variant: A comparison for risk for hospitalization and case-mortality rate for younger and older patients

PrgmNr 2320 Time to mortality and time to discharge from the hospital in Covid-19 data: A Turkish study

PrgmNr 2321 TransformerPRS: a deep learning-based polygenic risk model using bidirectional transformers derived from a language model

PrgmNr 2322 Using a digital health tool to analyse UK primary care electronic health records to identify potential cases of undiagnosed Marfan syndrome

PrgmNr 2323 A machine-learning-based prediction of disease endpoints from administrative health data accounts for misclassification and disease liability in GWAS

PrgmNr 2324 Characterization of the human ABO genotypes and their association to inflammatory and cardiovascular disease

PrgmNr 2325 Effects of GLP1R coding variation on random glucose levels and functional responses to GLP-1R agonists using whole exome sequencing data

PrgmNr 2326 Genetic and environmental determinants of drug adherence and drug purchasing behaviour

PrgmNr 2327 Higher adisposity, depression and wellbeing: causal inference using Mendelian randomization

PrgmNr 2328 HLA Fine Mapping and Genetic Risk Factors in Tuberculosis

PrgmNr 2329 Secondary findings from exome sequencing of 15,000 South Asians

PrgmNr 2330 Separating the effects of cardiometabolic risk factors from type 2 diabetes on coronary and peripheral artery disease

PrgmNr 2331 Transcriptome and GWAS identified rs2238678 as a critical SNP for eosinophilia in primary biliary cholangitis

PrgmNr 2332 Development of a novel, instrument-free, single-cell RNA sequencing technology (PIPseq) and its application to drug pathway discovery in lung cancer

PrgmNr 2333 Diversity of germline mutations in a cohort of unselected Brazilian pancreatic carcinoma patients

PrgmNr 2334 Effect of return of unknown significance on surgical variant of decision making in women with invasive breast cancer

PrgmNr 2335 Exome sequencing of 500 Brazilian patients with rare diseases: what we have learned

PrgmNr 2336 Human genetics isgreat:Incidental karyotype finding of aBRCA1translocation explaining family cancer history

PrgmNr 2337 Searching for epigenetic markers with prognostic value in Acute Myeloblastic Leukemia patients of Uruguay

PrgmNr 2338 Translocation t(8;14)(q11;q32) in Leukemia, a literature review

PrgmNr 2339 BRCA1 and BRCA2 variants of unclear significance in hereditary breast and ovarian cancer patients. Looking for novel pathogenic variants

PrgmNr 2340 Breast and prostate cancer risks for male BRCA1 and BRCA2 pathogenic variant carriers using recently-developed polygenic risk scores

PrgmNr 2341 Breast cancer risks associated with missense variants in breast cancer susceptibility genes

PrgmNr 2342 Copy number variation (CNV) analysis identifies variants in 1p36 in African American and Caucasian hereditary prostate cancer cases

PrgmNr 2343 Genetically Predicted Serum Insulin-like Growth Factor-1 (IGF-1) Levels and Risk of Multiple Primary Cancers in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial

PrgmNr 2344 Germline PALB2 mutations found in NGS multigenetic panels from breast-ovarian cancer patients in Argentina

PrgmNr 2345 Germline genetically predicted mammographic density is associated with breast tumor transcriptomic and proteomic features, CD8+ T cell infiltration, somatic mutational signatures, and tumor mutational burden in The Cancer Genome Atlas

PrgmNr 2346 Identification of copy number variants in hereditary lung cancer families

PrgmNr 2347 Immune-related intra-tumoral gene expression and survival by BRCA1/2 mutation status in high grade serous ovarian cancer cases: an Ovarian Tumor Tissue Analysis (OTTA) study

PrgmNr 2348 Malignant pleural mesothelioma: germline variants may steer tailored treatment

PrgmNr 2349 On cross-ancestry cancer polygenic risk scores

PrgmNr 2350 Rates of mutations in DNA repair genes are low in localized prostate cancer patients, particularly among non-White individuals

PrgmNr 2351 The IMPACT study Lynch Syndrome cohort - results after first round of screening

PrgmNr 2352 Wood stove use interacts with genetic variants to alter risk of gastric cancer in a high altitude Central American population

PrgmNr 2353 ACSL VI:A Novel Variation Results in Structural Modification and Multiple Neoplasia in a 46-year-old Female

PrgmNr 2355 Functional study of the 2q22 renal cancer susceptibility locus

PrgmNr 2356 Inflammatory and oncological processes contributing to the increased susceptibility to Barrett's esophagus in adults born with esophageal atresia

PrgmNr 2358 Utility of altered nuclear expression of the BAP1 for assessment of pathogenicity of Variants of Unknown Significance

PrgmNr 2359 Chromatin accessibility mapping in T-cell Prolymphocytic Leukemia

PrgmNr 2360 DNA Methylation Profiles of Ovarian Clear Cell Carcinoma

PrgmNr 2361 From genetics to immune system: the dynamic crosstalk in a wide cohort of completely resected Gastrointestinal Stromal Tumors (GISTs)

PrgmNr 2362 Genome Variation Analysis on Accelerated Framework

PrgmNr 2364 Germline variants associate with the presence of somatic TP53 and PIK3CA mutations in breast tumors

PrgmNr 2365 Identify non-mutational TP53 loss of function in human cancers

PrgmNr 2366 Integrated Transcriptomic Profiling with Deep Learning Classification Applied to Prostate Cancer Tissue Images

PrgmNr 2367 Integrative genomic analysis reveals new insights on complex mutational landscape in T-cell prolymphocytic leukemia

PrgmNr 2368 Investigating associations of common ATM missense variants with neoplasms in the UK Biobank exome sequencing data - a conditional analysis

PrgmNr 2369 JBrowse 2: an extensible open-source platform for modern genome analysis

PrgmNr 2370 Oncogenic mutations in the normal human brain

PrgmNr 2371 Paired DNA and RNA genetic testing improves variant detection and classification independently of personal history of cancer

PrgmNr 2372 Role of circulating immune checkpoints in KIT-mutated metastatic gastrointestinal stromal tumor (GIST) patients

PrgmNr 2373 Single cell consensus clustering (SC3) with REVEALTM: SingleCell on neuroblastoma and fetal gene expression data

PrgmNr 2374 Targeted NGS-based molecular profiling of conjunctival squamous cell carcinomas

PrgmNr 2375 The landscape and clinical relevance of somatic mutations in a cohort of patients with myelodysplastic syndrome

PrgmNr 2377 Using single-cell transcriptomics to fit a multi-level computational model of cancer

PrgmNr 2379 WGTS insights in a cancer patient: a tale of two primaries

PrgmNr 2380 Whole-genome sequencing of phenotypically distinct inflammatory breast cancers reveals similar genomic alterations to non-inflammatory breast cancers

PrgmNr 2381 A phenome-wide association scan reveals novel associations between genetically increased blood lipid levels and the risk of cholelithiasis

PrgmNr 2382 Assessing the contribution of rare variants to human complex traits by whole-genome sequencing

PrgmNr 2383 Assessing the role of rare pathogenic variants in heart failure progression in the UK Biobank

PrgmNr 2384 Chromosome X Inactivation is associated with cardiovascular risk and monocyte abundance in an ageing cohort

PrgmNr 2385 Discovery of rare variants associated with resting heart rate

PrgmNr 2386 Epigenetics of Single and Multisite Atherosclerosis in African Americans from the Genetic Epidemiology Network of Arteriopathy (GENOA)

PrgmNr 2387 Estimating the causal effects of cardiometabolic factors on coronary artery disease in British Pakistanis and Bangladeshis: A trans-ancestry Mendelian Randomisation study

PrgmNr 2388 Exploring causal relationships between cardiometabolic traits and cancers using pathway-specific genetic instruments

PrgmNr 2389 Gene-SCOUT: gene-based biomarker signatures can assist identification of novel genes from phenome-wide association analyses

PrgmNr 2391 Polygenic risk and its interaction with lifestyle for cardiovascular mortality

PrgmNr 2392 Polygenic score from a large GWAS predicts cases of heart failure with reduced ejection fraction (HFrEF) but not preserved ejection fraction (HFpEF)

PrgmNr 2393 The impact of polyunsaturated fatty acids biosynthesis on the risk of cardiovascular diseases: A Mendelian randomization study in up to 1,153,768 European ancestry individuals

PrgmNr 2394 The transcriptional landscape of human atria and its role in atrial fibrillation: the RACE-V Consortium

PrgmNr 2395 A glomerular transcriptomic landscape of APOL1 in Black patients with focal segmental glomerulosclerosis

PrgmNr 2396 Assessment of the Genetic Risk Scoring Algorithm GRS2 for Type 1 Diabetes in African American Children

PrgmNr 2397 Causal association between circulating metabolites and age at menarche: A two-sample Mendelian randomization study

PrgmNr 2398 Discovery of Type 2 Diabetes genes using an accessible tissue

PrgmNr 2399 Evaluation of genetic prediction and differential expression of genes in non-alcoholic fatty liver disease (NAFLD)

PrgmNr 2400 Gene-based analyses in population based data identify MTTP as associated with nonalcoholic fatty liver disease

PrgmNr 2401 Genetics of osteopontin in patients with chronic kidney disease: the German Chronic Kidney Disease study

PrgmNr 2403 Mirror effects of OPRD1 variants on diabetes and obesity

PrgmNr 2404 Mutation Burden Analysis Reveal Genetic Differences between Obesity and non-Obesity Asthma patients in African American Population

PrgmNr 2405 N6-methyladenosine methylation regulatory genes modified the effect of APOL1 risk genotype on progression of chronic kidney disease

PrgmNr 2406 Random glucose GWAS in 493,036 individuals provides evidence for reduced lung function as a novel complication of type 2 diabetes

PrgmNr 2407 The association of fasting glucose with sweetened beverages consumption is greater in Latin Americans with a high polygenic risk score for type 2 diabetes mellitus

PrgmNr 2408 The genetic architecture of BMI in infancy and early childhood is characterized by age-specific effects and overlap with pathways involved in Mendelian obesity

PrgmNr 2410 Whole-exome sequencing study identifies rare and low-frequency coding variants associated with diabetic nephropathy

PrgmNr 2411 Defining the Regulatory Mechanism for Helios in Autoimmune Disease

PrgmNr 2412 Evaluate mitophagy/autophagy modulators and FDA-approved drugs to prevent or ameliorate the autoimmune, lipodystrophic and neurodegenerative phenotypes observed in Clec16a KO mice

PrgmNr 2413 Expanded genome-wide association study of IBD identifies novel loci and implicates new genes in disease susceptibility

PrgmNr 2414 Risk HLA alleles in South America and potential new epitopes for SARS-CoV2

PrgmNr 2415 SARS-CoV-2 as a trigger of autoimmune diseases: brazilian experience

PrgmNr 2416 Whole-exome sequencing of hundreds of microbiopsies reveals the somatic evolutionary landscape of psoriatic skin

PrgmNr 2417 A de novo paradigm for male infertility

PrgmNr 2418 Analysis of transcriptional changes associated with pubertal development

PrgmNr 2419 Construction of copy number variation map identifies small regions of overlap and candidate genes for atypical female genitalia

PrgmNr 2420 Genetic architecture of severe spermatogenic failure and male infertility: time to advance standard of care

PrgmNr 2421 Idiopathic hypogonadotropic hypogonadism in the Maltese island population: A spectrum of phenotypes and genotypes

PrgmNr 2422 Challenging diversity in precision medicine through large-scale exome sequencing and genome wide genotyping in South Asians: Analyses of quantitative red blood cell traits

PrgmNr 2423 Clonal myelopoiesis and risk of adverse events in chronic kidney disease

PrgmNr 2424 A Haptoglobin (HP) Exon Deletion Polymorphism Associates with Alzheimer's Disease Status with Stratification of APOEe4 Status

PrgmNr 2425 Deep learning-based feature extraction in neuroimaging genetics for Alzheimer's Disease

PrgmNr 2426 Evaluation of the role of common variant burden in the familial aggregation of epilepsy

PrgmNr 2427 Exploring genome wide association signals in mesial temporal lobe epilepsy

PrgmNr 2428 Investigating the contribution of common variants to neurodevelopmental disorders: Expression-modifying variants and polygenic scores

PrgmNr 2429 Mendelian Randomization analysis reveals a potential causal relationship for lipids in amyotrophic lateral sclerosis risk

PrgmNr 2430 Polygenic risk score in Multiple Sclerosis: understanding the challenges of the post-GWAS era for complex diseases

PrgmNr 2431 Whole genome sequencing association analysis of general cognitive function in a multi-ethnic sample from the Trans-Omics for Precision Medicine (TOPMed) Program

PrgmNr 2432 An IQ-matched genetic comparison between cases with ASD and typically-developing controls

PrgmNr 2433 Analysis of age- and sex-specific comorbidity patterns among brain disorders

PrgmNr 2434 Exome reanalysis of 242 Brazilian trios reveals NPAS3 as a new possible novel autism risk locus

PrgmNr 2435 Family-based whole-genome sequencing studies of psychiatric disorders in youth

PrgmNr 2436 Genome-wide rare variant score associates with morphological subtypes of Autism Spectrum Disorder

PrgmNr 2437 Integrating Brain Imaging Phenotypes, Genomics, and Substance Use

PrgmNr 2438 Interactions between variants in detoxification and permeability barrier genes and xenobiotics in Autism Spectrum Disorder

PrgmNr 2439 Polygenic scores for cortical and subcortical brain structures are associated with cognition

PrgmNr 2441 Trans-ancestry GWAS meta-analysis of alcohol and tobacco addiction in 3.4 million individuals

PrgmNr 2442 A genome-wide association study of Ménière disease in the UK Biobank cohort reveals one genome-wide significant locus and genetic correlation with mood/anxiety-related traits

PrgmNr 2443 A phenotype-driven machine-learning model to assess the pathogenicity of digenic variant combinations: a case study on ciliopathies

PrgmNr 2444 Associated anomalies in cases with oral clefts

PrgmNr 2445 Characteristics of keratoconus corneal epithelium in multi-omic approach

PrgmNr 2447 Extreme Expression Genotypes revealed regulatory effects on gene expression linked to traits and footprints of positive selection

PrgmNr 2448 From GWAS to GWANN: Genome Wide Artificial Neural Networks and the interpretability of non-coding associations

PrgmNr 2449 Fusion of multiplex whole-exome and low-depth whole-genome sequencing for cost-effective rare variant assessment in the genome-wide association studies

PrgmNr 2450 Genetically independent phenotype analysis identifies LPA and VCAM1 as drug targets for human ageing

PrgmNr 2451 Genome-wide investigation of the interplay between age-related hearing loss, tobacco smoking, and noise

PrgmNr 2452 Global Biobank Meta-Analysis Initiative: A genome-wide association meta-analysis identifies novel primary open-angle glaucoma loci and shared biology with vascular mechanisms

PrgmNr 2453 GWAS and ExWAS of blood Mitochondrial DNA copy number identifies 71 loci and highlights a potential causal role in dementia

PrgmNr 2454 Higher BMI causes lower odds of depression in individuals of East Asian Ancestry

PrgmNr 2455 Identification of known and novel long non-coding RNAs potentially responsible for the effects of BMD GWAS loci

PrgmNr 2456 Identifying the mediator metabolites between obesity and osteoporosis through high-dimensional mediation analysis

PrgmNr 2457 Incorporating the gene-gene regulatory network into the estimation of effect size

PrgmNr 2458 Mapping the proteo-genomic convergence of human diseases

PrgmNr 2459 Mitochondrial and sex chromosome genetically regulated gene expression implicates new genes in complex traits across multiple human populations

PrgmNr 2460 Parent-of-origin inference in the UK biobank reveals new imprinted variants

PrgmNr 2461 Phenome-wide association study of clonal haematopoiesis somatic mutations in 391,756 UK Biobank participants

PrgmNr 2462 Rare, pathogenic variants causing short stature show incomplete penetrance in the UK Biobank

PrgmNr 2463 Shared genetic effects for type 2 diabetes and four cancers uncovered through multi-phenotype genome-wide association study

PrgmNr 2464 The chromatin accessibility signature of selection for exercise capacity in skeletal muscle

PrgmNr 2465 The effect of mitochondrial population genome variation on nuclear gene expression and its link to disease

PrgmNr 2466 The Netherlands Neurogenomics Database - Genetic susceptibility for brain disorders, clinical disease trajectories and end-stage neuropathology

PrgmNr 2467 Trio clinical exome analysis reveals a digenic phenotype in a 19-months-old boy with congenital deafness

PrgmNr 2468 Validation of cross-ancestry and cross-trait polygenic risk scores in the Genomics England cohort

PrgmNr 2469 Whole genome sequencing association study among 24,870 individuals from multiple ancestry-ethnicity groups identifies common and rare variants associated with serum urate

PrgmNr 2470 Whole-exome association analyses of polygenic deviators for height in 183,948 individuals from the UK Biobank

PrgmNr 2471 PAX6-associated microphthalmia: Transcriptome-wide investigation of patient-derived optic vesicle iPSC models reveals molecular disruption during early eye development associated with missense PAX6 variant

PrgmNr 2472 A Novel Method of Cell-type-level EWAS of HIV Infection Following Deconvoluting DNA Methylation in Blood

PrgmNr 2473 Association of peripheral blood microRNAs with cardiovascular endophenotypes in a population-based study

PrgmNr 2474 Characterization of caffeine response regulatory variants in vascular endothelial cells

PrgmNr 2475 Converging evidence for differential regulatory control of APOEε4 on African versus European haplotypes

PrgmNr 2476 Genetic Variation in Cis Regulatory Domains and Trans Regulatory Hubs of Immunity

PrgmNr 2477 Insulin Resistance is associated with PPARGC1A promoter DNA-methylation in Ecuadorian women with Turner syndrome

PrgmNr 2478 Investigating the role of a cardiac enhancer in the development of heart failure

PrgmNr 2479 Linking GWAS signals to target genes in skeletal muscle using single-nuclei multi-omic analyses

PrgmNr 2480 New epigenomic panel to detect mendelian disorders in Brazilian health system

PrgmNr 2481 Promoter methylation analysis in candidate genes of severe preterm birth

PrgmNr 2482 Using cellular deconvolution to investigate cell subtype proportions in cortical gene expression data in schizophrenia

PrgmNr 2483 Cas9 targeted enrichment of mobile elements using nanopore sequencing

PrgmNr 2484 Comparative transcriptomics of the human lung and epididymis epithelia

PrgmNr 2485 Detecting differentially methylated regions in individual methylomes

PrgmNr 2486 Direct assessment of impact of transcription factors on enhancer activity

PrgmNr 2488 Escape from X-inactivation in twins exhibits intra- and inter-individual variability across tissues

PrgmNr 2489 Genetic variants regulating microRNA expression in the developing human neocortex

PrgmNr 2490 Mapping chromatin interactions within the CYP3A gene locus

PrgmNr 2491 Muscle single nuclei association of cell type proportions, cell type chromatin accessibility and gene expression with sex and T2D-related physiological traits

PrgmNr 2492 Quantifying the shared genetic effects on the regulation of expression and protein levels in related individuals

PrgmNr 2493 Sex-specific and Generational Effects of Alcohol Use and Tobacco Smoking on Epigenetic Age Acceleration in the Michigan Longitudinal Study

PrgmNr 2494 Stressing out epigenetics: How chronic stress alters the epigenetic landscape during neurodevelopment

PrgmNr 2495 Transcriptome profiling using long-read sequencing to dissect the interplay between genetic variant and transcript variations in a population-based study

PrgmNr 2496 Transcriptomic mapping of bursting kinetics in human cells

PrgmNr 2497 Chromolooper:Building genetic neighborhoods using a physically informed graph algorithm

PrgmNr 2498 Differential network analysis of time-series RNA-sequencing in response to perturbation

PrgmNr 2499 Effective gene expression prediction from sequence by integrating long-range interactions

PrgmNr 2500 Individual Specific Networks Hot Zone Detection

PrgmNr 2501 Local gene co-expression: prevalence and molecular features across human tissues and single-cells

PrgmNr 2502 Modeling gene expression on the X chromosome for transcriptome-wide association studies

PrgmNr 2503 A signature of Neanderthal introgression on molecular mechanisms of environmental responses

PrgmNr 2504 Allelic frequency comparison to improve epidemiological decisions on hereditary diseases in a Latino population

PrgmNr 2507 Characterizing the evolutionary history of HLA genes in sub-Saharan African populations

PrgmNr 2508 Deep variational autoencoders for population genetics: applications in classification, imputation, dimensionality reduction, and novel lossless data compression

PrgmNr 2509 Estimating ancestry-specific allele frequencies in admixed samples from large-scale sequencing data

PrgmNr 2510 Exploring the role of local sequence context on patterns of germline mutation

PrgmNr 2511 Genetic and pharmacological causes of germline hypermutation

PrgmNr 2512 Genetic structure, kinship, and ancient pathogen diversity in a Moche tomb in Northern Peru

PrgmNr 2513 Population genetic analyses implicate biogenesis of translation machinery in human ageing

PrgmNr 2514 Revealing the recent demographic history of Europe via haplotype sharing in the UK Biobank

PrgmNr 2515 Sequencing the general population at a fine-geographic scale: the French POPGEN project

PrgmNr 2516 SIRPA V1/V2 haplotypes were selected by balancing selection

PrgmNr 2517 The genetic landscape of South Native American populations

PrgmNr 2518 The GLAD Initiative: a database for supporting population genetics and association studies in the Americas

PrgmNr 2519 Genetics Adviser: The development, usability and acceptance testing of a patient-centered digital health application to support clinical genomic testing

PrgmNr 2520 Building Genomic Data Science Capacity for Health Discovery and Innovation in Africa

PrgmNr 2521 Challenges to develop TBL modules that integrates genetics and genomics with basic and clinical disciplines in the first year of a Medical School

PrgmNr 2522 Consumer-initiated genetic testing: A look at the current landscape in the United States

PrgmNr 2523 Development and evaluation of digital decision aids with diverse patients: a systematic review

PrgmNr 2524 Divergent perceptions of meaning and utility of genetic testing results between families, clinicians, and laboratory directors

PrgmNr 2525 Engagement in science: virtual interaction as a flexible tool

PrgmNr 2526 Free digital genomic educational resources: evaluation of their international value and uptake

PrgmNr 2527 Including children in a polygenic risk score return of results pragmatic trial: Considerations and decisions by consensus in the eMERGE Network

PrgmNr 2528 Integration of Genetic Counseling Services into an Immunodeficiency Clinic: Roles of the Genetic Counselor and Impact on Patient Evaluation and Care

PrgmNr 2529 International education, awareness, celebration, and commemoration around the bicentennial of Gregor Mendel's birth

PrgmNr 2530 Participant Engagement and Cancer Genome Sequencing (PE-CGS) Network

PrgmNr 2531 Public Health Genetics Week 2021: A Public Awareness Campaign

PrgmNr 2532 RNA-Seq and recurrence risk testing for breast cancer in the context of an undergraduate course: Implications for patient education

PrgmNr 2533 The role of general practitioner in management of rare diseases

PrgmNr 2534 Transitioning to a virtual training model for building capacity in genomic analysis skills across world regions

PrgmNr 2535 User-centered design of genetic test reports leads to improvements on self-reported and objective measures

PrgmNr 2536 Efficient delivery of FMR1 across the blood brain barrier using AAVphp construct - feasibility of gene therapy for fragile X syndrome

PrgmNr 2537 Exon specific U1 snRNA therapeutic strategy to rescue retinal degeneration in familial dysautonomia

PrgmNr 2538 Nanoparticle-based non-viral CRISPR/Cas9 delivery platform for targeting thoracic aortic aneurysm associated heterozygous mutations

PrgmNr 2539 Additional data obtained from Exome/Genome Sequencing : two national studies to discuss the risk-benefit balance for implementation in France

PrgmNr 2540 Enhanced diagnosis mapping and exome sequencing identify novel variants for rare diseases in the UK Biobank

PrgmNr 2541 Generation and quality control of the biobank-scale methylation dataset in the VA's Million Veteran Program (MVP)

PrgmNr 2542 GHGA - The German Human Genome-Phenome Archive - A new federated national infrastructure for sharing human genomes in a European framework

PrgmNr 2543 Natural-language processing and machine learning can systematically infer tissue annotations of omics samples by modeling their unstructured metadata

PrgmNr 2544 NDD-CNV Portal: Facilitating genetic test interpretation, research and education for Neurodevelopmental CNVs

PrgmNr 2545 SCN Portal for epilepsy genetics: Facilitating genetic test interpretation, research and education for sodium channel disorders

PrgmNr 2546 Strategies to increase the availability of summary statistics in the GWAS Catalog

PrgmNr 2547 The Brazilian Rare Genomes Project: the largest whole genome sequencing effort for rare disorders in South America

PrgmNr 2548 Towards robust clinical genome interpretation: developing a consistent terminology to characterise disease-gene relationships for harmonised curation - an updated structural variant ontology for the NGS era

PrgmNr 2549 De novo pathogenic variants predominate in exomes of a cohort of patients with intellectual disabilities and/or congenital anomalies. Preliminary results from a pilot program for Rare Undiagnosed Diseases in Chile

PrgmNr 2550 De novo variants in TCF4 with a suspected gain-of-function mechanism are responsible for a new malformative disease without intellectual disability

PrgmNr 2551 Biallelic loss-of-function variants in CACHD1, encoding an α2δ-like voltage-gated calcium channel regulator, cause a neurodevelopmental, craniofacial, and genitourinary syndrome

PrgmNr 2552 NSD1 gene variations associated to Sotos syndrome in humans and diverse selection within exons

PrgmNr 2553 Clinical delineation, sex differences and genotype-phenotype correlation in pathogenic KDM6A variants causing X-linked Kabuki syndrome Type 2

PrgmNr 2554 Diverse phenotype genotype characterization and novel SOX9 variation in Campomelic Dysplasia

PrgmNr 2555 Exome sequencing identifies a missense mutation in MMP14 in autosomal dominant hypotrichosis

PrgmNr 2556 First identification of an individual with atypical Smith-Magenis syndrome without intellectual disability and non-coding structural variant

PrgmNr 2558 Genotype-phenotype correlations in PIK3CA-related overgrowth spectrum (PROS) and overlapping phenotypes: a systematic review of 1007 patients with PIK3CA pathogenic variants

PrgmNr 2559 Germline 3q21 Chromosome Deletion: A Rare Multi-System Condition

PrgmNr 2560 Neuropsychiatric and behavourial phenotype description of 24 Brazilian patients harboring Cornelia de Lange Syndrome

PrgmNr 2561 Sequencing analysis of genes involved in the Wnt signaling pathways reveals novel candidates for nonsyndromic cleft lip with or without cleft palate

PrgmNr 2562 A homozygous missense variant in amyloid-beta precursor protein (APP) may cause severe syndromic intellectual disability

PrgmNr 2563 A nonsense EBF3 variant in an individual with intellectual disability, without ataxia and hypotonia

PrgmNr 2564 A predictive modeling strategy to explore the mechanism(s) underlying dominantly-inherited peripheral neuropathy caused by aminoacyl-tRNA synthetase variants

PrgmNr 2565 Biallelic frameshift variant in the TBC1D2B gene in two siblings with progressive cognitive impairment, gingival overgrowth, limb tremor, and fribous dysplasia of face

PrgmNr 2566 Biallelic SEPSECS variants in two siblings with pontocerebellar hypoplasia type 2D underscore the relevance of splice-disrupting synonymous variants in disease

PrgmNr 2567 CAPRIN1: how a recurrent de novo mutation causes a progressive early onset neurodegenerative disorder in a prion-like domain-harboring protein

PrgmNr 2568 Characterization of Phelan-McDermid brazilian cohort with small deletions and point mutations: focus on regression and behavioral changes across different ages

PrgmNr 2569 Characterization of the GARS1 mRNA interactome in glycyl-tRNA synthetase (GARS1)-related neurological disease

PrgmNr 2570 Complex inheritance in spastic paraplegia 7

PrgmNr 2571 Defining the molecular mechanism ofGARS1-related Charcot-Marie-Tooth disease

PrgmNr 2572 Defining the neurodevelopmental spectrum of GABRA5-associated disease

PrgmNr 2573 Elucidating genomic and transcriptomic hexanucleotide repeat number and methylation in the X-linked dystonia-parkinsonism-relevant retrotransposon insertion by Nanopore sequencing

PrgmNr 2574 Exome sequencing in 200 intellectual disability/autistic patients: new candidates and atypical presentations

PrgmNr 2575 Exploring the phenotypic spectrum of ARID1B-related disorders: Further two case reports

PrgmNr 2576 Factors influencing the chance of a genetic diagnosis in >13,000 individuals with developmental disorders

PrgmNr 2577 Human patient SFPQ homozygous mutation is found deleterious for brain and motor development in a zebrafish model

PrgmNr 2578 Identification of rare copy number variants in a family with high prevalence of psychosis

PrgmNr 2579 Infantile ascending spastic paraplegia linked to ALS2pathogenic variant

PrgmNr 2580 Inherited variants in CHD3 demonstrate variable expressivity in Snijders Blok-Campeau syndrome

PrgmNr 2581 Methylation status of chromosome 5 in Cri du ChatSyndrome: new insights for clinical variability?

PrgmNr 2582 Modeling a neuropathy-associated GARS1 mutation in C. elegans

PrgmNr 2583 Monoallelic variants in TFAP2E are implicated in craniofacial and central nervous system anomalies

PrgmNr 2584 NeuroWES-Macedonia: genetic diagnoses for complex syndromic cases with neurodevelopmental disorders

PrgmNr 2585 Paediatric neurogenetic disorders in Bangladesh: First tier exome sequencing in previously unexplored population

PrgmNr 2586 Phenotypic characteristics of a family with distal arthrogryposis type 2B (Sheldon-Hall syndrome) due toTNNT3pathogenic variant

PrgmNr 2587 Recessive burden analysis and gene discovery in ~40,000 developmental disorder trios with diverse ancestries

PrgmNr 2588 Recessive inactivating variants in DMAP1 in an intellectual disability syndrome

PrgmNr 2589 The phenotypic spectrum of BRAT1-related disorders : 51 additional cases and literature review

PrgmNr 2590 The pleiotropy of neurodegenerative repeat expansions in ALS

PrgmNr 2591 Whole exome sequencing of consanguineous individuals reveals rare coding variants in known and novel genes that contribute to neurological disorders

PrgmNr 2592 Clinical implementation of RNA sequencing for Mendelian disease diagnostics

PrgmNr 2593 Diagnosing pediatric mitochondrial disease: Lessons from 2,000 exomes

PrgmNr 2594 Dietary nutritional recommendations in patients with phenylketonuria in Ecuador

PrgmNr 2595 Dysosteosclerosis in an individual harboring a homozygous nonsense variant in TNFRSF11A: further considerations on genotype-phenotype correlation

PrgmNr 2596 Expanding the clinical spectrum of NDUFAF3-related disorders: Prolonged survival and unusual course in a patient with compound heterozygous NDUFAF3 variants

PrgmNr 2597 Gut microbiome affects human bone mass, microarchitecture, and strength: The Framingham Study

PrgmNr 2598 Novel RUNX2 gene variation associated with cleidocranial dysplasia-like phenotype

PrgmNr 2599 Novel variants of PAH gene in Ecuadorian phenylketonuria patients: Do we need new molecular panels for metabolic disorders for Latin American populations?

PrgmNr 2600 Sensorineural hearing loss is an early indicator of mitochondrial disease

PrgmNr 2601 The role of seryl-tRNA synthetase in mitochondrial biology and disease

PrgmNr 2602 A digital health approach to rare disease diagnosis; identifying patients with feature combinations warranting further evaluation for a possible diagnosis of 22q11.2 deletion syndrome from UK primary care electronic health records

PrgmNr 2603 Association of missense variants in the transthyretin TTR gene with hereditary transthyretin-mediated hATTR amyloidosis in 454,782 UK Biobank whole exome sequences

PrgmNr 2604 Genetic Overlap of Acute Myocarditis and Inherited Cardiomyopathy

PrgmNr 2605 Increasing the diagnostic yield through post-report reanalysis of exome data in subjects with immune disorders

PrgmNr 2606 Pediatric onset cardiomyopathy secondary to TNNC1 heterozygous and homozygous mutations: a French cohort study

PrgmNr 2607 Rare variant burden influences the rate of disease progression in Polycystic Kidney Disease

PrgmNr 2608 13q Deletion Syndrome involving RB1: characterization of a new minimal critical region for psychomotor delay

PrgmNr 2609 7q36.1q36.2 deletion in a child with intellectual disability, developmental delay, short stature, and a Chiari 1 malformation

PrgmNr 2610 Diagnostic yield of exome sequencing in neurodevelopmental disorders in a Brazilian clinical laboratory

PrgmNr 2611 Genetic evaluation of children born with very low birth weight

PrgmNr 2612 Lung Pathology in Takenouchi-Kosaki Syndrome: Expanding the phenotypic spectrum

PrgmNr 2613 Missense variants at a conserved 14-3-3 binding site in HDAC4 result in a novel intellectual disability syndrome

PrgmNr 2614 Molecular and clinical diagnosis of 17p11.2 microduplication described as Potocki-Lupski Syndrome: Case Report

PrgmNr 2615 Automated variant classification workflows of known and novel variants can be used to maintain quality standards, support standardization and reduce turn-around times in a rare disease laboratory

PrgmNr 2616 Brazil enters the genome sequencing era: The Rare Genomes Project - interim results

PrgmNr 2617 Clinical relevance of the study of pigmentary mosaicism using genomic array

PrgmNr 2618 Fold-back mechanism originating an inv dup del rearrangement in chromosome 15

PrgmNr 2620 Incorporating a risk modifying SNP (rs143838139) to increase the detection of [2+0] carriers for Spinal Muscular Atrophy

PrgmNr 2622 The ACMG SF v3.0 gene list increases actionable variant detection by 25% over v2.0 in the ClinSeq® cohort

PrgmNr 2623 The knowns and unknowns: Challenges of Genes of Uncertain Significance in the era of high-throughput sequencing and lessons from Matching Tools

PrgmNr 2624 What does a 45,X karyotype add to 5p- syndrome literature?

PrgmNr 2625 APOE4 increases astrocyte reactivity in cells with local European ancestry

PrgmNr 2626 A murine model lacking Lyst displays progressive neurodegeneration and recapitulates Chediak-Higashi Syndrome

PrgmNr 2627 A splice site SNP in Gabra2 is a modifier of Scn8a encephalopathy in the mouse

PrgmNr 2628 Characterization of the genetic architecture of neurodevelopmental disorders using Duchenne Muscular Dystrophy as a model

PrgmNr 2629 Effects of somatic mutations on cellular differentiation outcomes and outlier behaviour in iPSC models
PrgmNr 2630 Pathogenic SPTBN1 variants cause an autosomal dominant neurodevelopmental syndrome
PrgmNr 2631 Schizophrenia risk mediated by microRNA target genes in 22q11.2 deletion syndrome
PrgmNr 2632 Telomere length analysis in amyotrophic lateral sclerosis using large-scale whole genome sequence data
PrgmNr 2633 Understanding the role of CLP1 in mammalian mRNA transcription and cleavage in neurodegeneration
PrgmNr 2634 Wnt/beta-catenin pathway and cell adhesion deregulation in CSDE1-related intellectual disability and autism spectrum disorders
PrgmNr 2635 An epigenome-wide view of osteoarthritis in primary tissues
PrgmNr 2636 Concurrence of Incontinentia pigmenti and thrombophilia increases the risk of Recurrent Early Pregnancy Loss
PrgmNr 2637 Neonatal onset and severe complicated phenotype caused by heterozygous or homozygous new ATL1 gene mutations
PrgmNr 2638 Perturbation in hepatic lipid metabolism associated with altered TM6SF2 gene expression in HuH-7 cells
PrgmNr 2639 Phenotypic Consequences of Variation in Elastin
PrgmNr 2640 Saturation-scale functional evidence supports clinical variant interpretation in Lynch Syndrome
PrgmNr 2642 Spatial analyses of complement factor genes in AMD: Clarifying known genetic relationships and creating prediction models
PrgmNr 2643 A network-based deep learning framework translates GWAS and multi-genomics findings to pathobiology and drug discovery for Alzheimer’s disease
PrgmNr 2644 A pipeline for the identification of rare variants in multigenerational families with high prevalence of complex disorders
PrgmNr 2645 Aberrant splicing prediction across human tissues
PrgmNr 2646 Analysis of 5’UTR Variation in Rare Disease Patients Reveals Variants of Potential Disease Relevance
PrgmNr 2647 Characterization of complex genomic rearrangements detected with long reads sequencing: two chromothripsis cases
PrgmNr 2648 Improved detection of allele-specific expression using Personalised ASE Caller (PAC)
PrgmNr 2649 MRSD: a novel quantitative approach to predict tissue-specific feasibility of RNA-seq in the investigation of splicing diversity
PrgmNr 2650 Re-genotyping and tool combinations improve detection accuracy of Copy Number Variants in Whole Genome Sequencing data
PrgmNr 2651 A large-scale cohort metagenomic study of the Brazilian miscegenated population: building gut microbiome profiles of healthy and unhealthy individuals
PrgmNr 2653 High yield depletion for an RNA-seq application using an innovative benchtop platform for workflow automation and miniaturization
PrgmNr 2654 Integrating expression and methylation data into GWAS for the discovery of novel colorectal cancer risk loci
PrgmNr 2655 Lipid pathway signatures discriminate subgroups linked to race/ethnicity in metabolomic and proteomic analysis from the TOPMed Study
PrgmNr 2656 MED13L knockout in cerebral organoids disrupts retinal gene expression regulation
PrgmNr 2657 Same role but different actors in the genetic regulation of post-translational modification of two distinct proteins
PrgmNr 2658 Understanding the genetic contribution to neuroinflammatory cell traits using iPSC-derived microglia
PrgmNr 2659 Whole Blood RNA Sequencing in a Cohort of Undiagnosed Pediatric Patients
PrgmNr 2660 de novo variant calling identifies cancer mutation profiles in the 1000 Genomes Project
PrgmNr 2661 Building a predictive matrix model for autism genotype-phenotype associations

PrgmNr 2662 CoverageMaster: a clinical grade and user oriented copy number variant caller

PrgmNr 2663 Highly sensitive transcriptomic-based pooled CRISPR screening on 1 million cells enabled by spatial molecular imager

PrgmNr 2664 Longitudinal multiomics data responses aid the detection of collective phenotypic characteristics in prediabetic/diabetic monitoring

PrgmNr 2665 Meta-Learning on Next-Generation Sequencing data to Improve Cell Type Annotation

PrgmNr 2667 OUTRIDER2: A generalized framework for context-dependent outlier detection in omics data

PrgmNr 2668 The value of long read genome sequencing in the diagnostic evaluation of muscular dystrophy

PrgmNr 2669 Using deep learning to detect Noonan syndrome from electronic health records

PrgmNr 2670 Using transcriptomic analysis to identify pathogenic variants in rare Mendelian diseases missed on conventional genomic testing

PrgmNr 2671 Clinical pharmacogenomic test reporting preferences among adult and pediatric providers

PrgmNr 2672 Common statin intolerance variants in ABCB1 and LILRB5 show synergistic effects in statin response: an observational study using electronic health records

PrgmNr 2673 DrugnomeAI: An exome-wide machine-learning framework for druggability prediction and ranking of drug target candidates by stochastic semi-supervised learning

PrgmNr 2674 Pharmacogenetics for precision cardiovascular therapy: a novel GRS for SLCO1B1 as a predictor of statin intolerance

PrgmNr 2675 Sanford Health Preemptive Population Genetic Screening - A Comparison of Military Veterans and non-Veterans

PrgmNr 2676 Transcriptomic effects of propranolol and primidone uncover important pathways for tremor reduction in essential tremor

PrgmNr 2677 An additionnal case of arthrogryposis with agenesis of the corpus callosum related to SCYL2 diagnosed in the prenatal period

PrgmNr 2678 Association between acute maternal infections during pregnancy and birth defects in Bogota and Cali, Colombia 2001-2018

PrgmNr 2679 Evidence for genetic imprinting and a causal role of fetal insulin and diastolic blood pressure in regulation of placental weight

PrgmNr 2680 Genome-wide genetic control of fetal placental genomics shows multiple associations with health and disease across the life course, informing the Developmental Origins of Health and Disease

PrgmNr 2681 Investigating the influence of fetal growth on maternal blood pressure in pregnancy

PrgmNr 2682 MD

PrgmNr 2683 Mining NIPT data of 108.349 Dutch pregnant women: cfDNA based viral analysis and GWAS

PrgmNr 2684 Small supernumerary marker chromosome 15 in Prenatal Diagnosis: A clinical challenge

PrgmNr 2685 A computationally efficient clustering linear combination approach to jointly analyze multiple phenotypes for GWAS

PrgmNr 2686 A method for estimating personalized causal genomic variants of disease

PrgmNr 2687 A powerful test of ancestral heterogeneity in the effects of gene expression on complex traits

PrgmNr 2688 An ensemble tool to predict health outcomes using microbiome data

PrgmNr 2689 An unsupervised approach to identifying haplogroups in mitochondrial DNAs

PrgmNr 2690 Combining the strengths of inverse-variance weighting and Egger regression in Mendelian randomization using a mixture of regressions model

PrgmNr 2691 Congenica A.I.: A High Scalable Machine Learning Decision Support Framework for Improving the Diagnostic Yield

PrgmNr 2692 Cost-effective Low-Coverage Sequencing Imputation from Millions of Haplotypes

PrgmNr 2693 Deep Learning in microbiome: Hybrid CNN-LSTM model for disease prediction using longitudinal microbiome data

PrgmNr 2694 Detecting differential isoform splicing from single-cell long-read RNA sequencing

PrgmNr 2695 Distinguishing different psychiatric disorders using polygenic prediction
PrgmNr 2696 Exploiting public GWAS databases to identify and adjust for heritable confounding in Mendelian randomization studies
PrgmNr 2697 Gene selection by incorporating genetic networks into case-control association studies
PrgmNr 2698 Genetic fine-mapping with dense linkage disequilibrium blocks
PrgmNr 2699 Identifying clinically-relevant circulating protein biomarkers for type 1 diabetes: A two-sample Mendelian randomization study
PrgmNr 2700 Incorporating family disease history and controlling case-control imbalance for population based genetic association studies
PrgmNr 2701 Inter-cohort heterogeneity significantly undermines fine-mapping a meta-analysis
PrgmNr 2702 Joint analysis model of multiple high-throughput sequencing data
PrgmNr 2703 Leveraging Machine Learning (ML) to Predict Inherited Variants Associated with Chronic Lymphocytic Leukemia (CLL)
PrgmNr 2704 Leveraging summary statistics fromEuropean GWAS to improve discovery of non European populations accounting for heterogeneity in effect sizes
PrgmNr 2705 M-DATA: A statistical approach to jointly analyzing de novo mutations
PrgmNr 2706 Modeling transcription factor binding in statistical tests of common and rare non-coding regulatory variants
PrgmNr 2707 Optimising Polygenic Score Portability Across Ancestral Populations
PrgmNr 2708 Ordinal regression with gene-based aggregation tests in the UK Biobank
PrgmNr 2709 PathWAS reveals biologically relevant pathways in complex traits
PrgmNr 2710 Penalized mediation models for multivariate data
PrgmNr 2711 Pleiotropy testing identifies genetic variants with pleiotropic evidence in Alzheimer's disease

PrgmNr 2713 scRegMap: A statistical framework for mapping context-specific regulatory variants using single cell RNA-sequencing
PrgmNr 2714 Simulated data provides insight on optimal control method for confounders
PrgmNr 2715 Simulations to understand potential miscalibration in fine-mapping algorithms
PrgmNr 2716 Spatially Aware Dimension Reduction for Spatial Transcriptomics
PrgmNr 2717 Spatially Informed Cell Type Deconvolution for Spatial Transcriptomics
PrgmNr 2718 Stacking Method to Improve Polygenic Risk Prediction in Admixed Individuals using Population Specific GWAS Effect Sizes
PrgmNr 2719 Testing Mendelian Randomization methods via zero-relevance analysis of COVID-19 traits
PrgmNr 2720 The construction of multi-ethnic polygenic risk score using transfer learning
PrgmNr 2721 The METRO method improves identification of gene-disease associations through multi-ancestry transcriptome-wide association studies
PrgmNr 2722 Ultrahigh dimensional learning of polygenic risk scores for Mendelian randomization studies

PrgmNr 2725 A phenome-wide association study of telomere length using combined qPCR and whole genome sequencing data in 136,597 UK Biobank participants
PrgmNr 2726 A tissue-level phenome-wide network map of colocalized genes and phenotypes in the UK Biobank
PrgmNr 2727 A tissue-specific allelic hierarchy predicts phenotypic outcomes for USH2A-related disorders in the RUSH2A study
PrgmNr 2728 Assisted reproductive technologies reduce fetal growth and alter maternal and fetal DNA methylation
PrgmNr 2729 Enigmas of TRAF3: Toward a mechanistic understanding of world differences in gallstone prevalence
PrgmNr 2730 Gene expression differences by race and genetic ancestry in B-cell acute lymphoblastic leukemia
PrgmNr 2731 Genetic Management in Benign Prostatic Hyperplasia

PrgmNr 2732 Identifying casual effects of lipids in neurological diseases using univariate and multivariate Mendelian Randomization
PrgmNr 2733 Investigating risk factors for perceived facial ageing: A Mendelian randomization study
PrgmNr 2734 Investigating the genetic interplay between adult height and disease
PrgmNr 2735 Large-scale genome-wide meta-analyses provide insights for the development of new disease modifying targets for osteoarthritis
PrgmNr 2736 Linkage analyses confirm chromosome 5q35 as a risk locus in African Ancestry population
PrgmNr 2737 Sex-specific genetic modifiers of hemolysis in blood donors
PrgmNr 2738 Unbiased metabolomics by Mendelian randomisation links serum isoleucine to risk of amyotrophic lateral sclerosis
PrgmNr 2739 Uncovering the role of genomic structural variations associated with risk of osteoporosis through whole genome sequencing
PrgmNr 2740 White matter integrity and nicotine dependence in smokers: evaluating vertical and horizontal pleiotropy
PrgmNr 2741 Amplification of CCND1/IGH@ Fusion in Plasma Cell Myeloma - A Case Report and Review of Literatures
PrgmNr 2742 Cancer Risk C, a functional genomics test, is a sensitive, specific, and rapid diagnostic for Lynch syndrome
PrgmNr 2743 Clinically relevant germline variants in a cohort of Mexican young women with breast cancer: analysis from whole-exome sequencing
PrgmNr 2744 Delivering patient-centered care in precision cancer medicine: A systematic review
PrgmNr 2745 Facilitators and barriers to implementing non-indication based genetic testing within corporate wellness programs
PrgmNr 2746 Genetic testing in patients with colorectal cancer; a real-world analysis of utilization and results
PrgmNr 2747 Prevalence of pathogenic mutations and variants of uncertain significance in Cambridge, MA
PrgmNr 2748 Tumor genomic profiling (TGP) - identified mutations in moderate risk breast and ovarian cancer genes: worthy targets for germline confirmatory testing
PrgmNr 2749 Disease risk prediction with two-phase study data
PrgmNr 2750 Evidence of divergent genetic associations across breast cancer molecular subtypes via distal mediator-enriched Transcriptome-Wide Association Study
PrgmNr 2751 Genetic susceptibility to smoking and alcohol use behaviors identifies differential risk in oropharyngeal and oral cavity cancer patients based on HPV status
PrgmNr 2752 High prevalence of subclinical micro-FAFs in facial skin in Tuberous Sclerosis Complex
PrgmNr 2753 PhenX presents a Variable Compare tool for COVID-19 and a new research domain: Cancer Outcomes and Survivorship
PrgmNr 2754 Prevalence of autoimmune diseases in Li-Fraumeni Syndrome
PrgmNr 2755 Prostate cancer susceptibility variants and their interplay with lifestyle factors on the risk of lethal disease
PrgmNr 2756 Quantification of cancer risk, patterns, and genotype-phenotype associations in Li-Fraumeni syndrome
PrgmNr 2757 USP8 gene may contribute to the development of bilateral adrenal hyperplasia and ACTH-independent Cushing syndrome
PrgmNr 2758 Cancer Risk B is a rapid and accurate test of peripheral blood mononuclear cells for prediction of breast cancer risk
PrgmNr 2759 Casting a Wide Net: Finding actionable results in non-breast cancer genes on multi-gene panel testing in a breast cancer cohort
PrgmNr 2760 Cellular diversity of human breast tissue evaluated in Pre-templated Instant Partitions for single-cell RNA sequencing (PIPseq)
PrgmNr 2761 Improving predictions of variant functional effects using deep mutational scanning data

PrgmNr 2762 Large-scale functional assessment of RAD51C missense variants with PARP inhibitor screens

PrgmNr 2763 Measuring (almost) all possibleSTK11/LKB1 missense variants in monogenic disease and cancer

PrgmNr 2764 Pathological variant c.8143C&gtT in ATM gene in a patient with non-Hodgkin lymphoma

PrgmNr 2765 Uncovering the functional variants and target genes of the 7q32 pancreatic cancer risk locus

PrgmNr 2766 3D confirmation analysis of the FGFR2 kinase domain with disease-associated mutations reveals novel features that correlate with clinical presentation

PrgmNr 2767 A genome-first approach characterizing the prevalence and cancer phenotype of pathogenic germline DICER1 variants in two large, unselected cohorts

PrgmNr 2768 An analysis of automated cell-type labelling algorithms for tumour micro-environment single-cell RNA sequencing data

PrgmNr 2769 Evaluating Immune Repertoire Patterns During Prostate Cancer Tumor Progression

PrgmNr 2770 Identification of the transcriptome signature of Beckwith-Wiedemann Syndrome hepatoblastoma predisposition

PrgmNr 2771 Inferring cancer evolution from single tumour biopsies using synthetic supervised learning

PrgmNr 2772 Long-read sequencing reveals alternative splicing of transposable elements in breast cancer

PrgmNr 2773 Multiple germline events lead to cancer development in patients with Li-Fraumeni syndrome

PrgmNr 2774 Optical genome mapping reveals novel structural variants in pediatric brain tumors

PrgmNr 2775 The putative tumor suppressor geneUSP13in familial papillary thyroid carcinoma

PrgmNr 2776 Treatment-specific gene regulatory pathways in colorectal cancer patients: CALGB/SWOG 80405 (Alliance)

PrgmNr 2777 Trichostatin A sensitivity signature across hematological cell lines

PrgmNr 2778 A meta-genetic risk score across cardiometabolic risk factors for cardiovascular mortality: a prospective cohort study of UK Biobank

PrgmNr 2779 Association of a novel locus with rheumatic heart disease in black African individuals: Findings from the RHDGen study

PrgmNr 2780 Cardiac enhancer variants at the MTSS1 locus are associated with smaller, more contractile hearts

PrgmNr 2781 Cell-type-specific gene expression and transcriptional networks reveal Adamts2 as a robust regulator of cardiac homeostasis during heart failure

PrgmNr 2782 Dissecting 90 percent of lipoprotein(a) heritability in 487,571 UK Biobank participants

PrgmNr 2783 Genetic association analysis of dilated cardiomyopathy in the VA Million Veteran Program

PrgmNr 2784 Genetic variation in/near LIPC, MGAT1, and APOA1 is associated with longitudinal change in serum lipid levels in Samoan adults

PrgmNr 2785 Genome-wide association study of coronary heart disease in chronic kidney disease: the Chronic Renal Insufficiency Cohort (CRIC) study

PrgmNr 2786 Leveraging a founder population to identify novel rare-population genetic determinants of lipidome

PrgmNr 2787 Multi-trait GWAS of atherosclerosis detects novel loci and potential therapeutic targets

PrgmNr 2788 Prioritizing genes from genome-wide meta-analyses of blood lipid levels

PrgmNr 2789 Proteomic analysis of 92 circulating proteins and their effects in cardio-metabolic diseases

PrgmNr 2790 Rare variant carrier status on risk of coronary artery disease among individuals with high polygenic risk

PrgmNr 2791 OCRL tissue-specific alternative splicing in humans results in divergent phenotypes

PrgmNr 2792 Adiponectin and cardiometabolic outcomes in sub-Saharan Africans: A Mendelian randomization study

PrgmNr 2793 Analyzing human knockouts to validate GPR151 as a therapeutic target for reduction of body mass index

PrgmNr 2794 Circulating metabolite biomarkers for osteoporosis: Evidence from Mendelian randomization and direct metabolite measurement

PrgmNr 2795 Constructing Metabolic Syndrome-related polygenic scores in the multiethnic GENNID family study

PrgmNr 2796 Enrichment of Motilin Receptor LOF variants in Gastroparesis

PrgmNr 2797 Exome-wide association analyses of MRI imaging-derived traits in up to 32k participants of the UK Biobank identify genes and variants involved in the regulation of body composition

PrgmNr 2798 G6PD deficiency associated genetic variants and the diagnostic utility of HbA1c: the Africans in America Study

PrgmNr 2799 Genome-wide assessments of obesity and 138 inflammatory markers to characterize obesity-induced inflammation

PrgmNr 2801 Integrating genetics and plasma proteomics to predict risk of type 2 diabetes in adults and children: A two-sample Mendelian randomization study

PrgmNr 2802 Integrative multiomics analyses reveals link between obesity and insulin resistance-associated plasma metabolites and adipose tissue transcripts in African Americans

PrgmNr 2803 Machine Learning for Metabolite Estimation to Examine Contributors to Adiposity: The GUARDIAN Consortium

PrgmNr 2804 Mapping genomic regulation of kidney diseases and traits at a cell type and variant level of specificity

PrgmNr 2805 Metabolomic signatures of waist circumference in the Hispanic Community Health Study/ Study of Latinos (HCHS/SOL)

PrgmNr 2806 Methylation profiles at birth linked to early childhood obesity

PrgmNr 2807 Multi-ethnic genome-wide meta-analysis in type 1 diabetes

PrgmNr 2808 Novel Genetic Loci Associated with Age-of-onset of Type 1 Diabetes

PrgmNr 2809 Pleiotropic Loci Influencing Cystic Fibrosis-Related Diabetes and Neonoatal Intestinal Obstruction

PrgmNr 2810 Polygenic susceptibility to obesity modifies the impact of salivary amylase gene (AMY1) copy number on BMI

PrgmNr 2811 Role of diet and submaximal cardiorespiratory fitness as mediators of genetic susceptibility to obesity: results from the Quebec Family Study

PrgmNr 2812 Therapeutic validation of FATP5/SLC27A5 using loss-of-function variation in humans

PrgmNr 2813 Trans-ancestry meta-analyses for waist-hip ratio in >1.1M individuals

PrgmNr 2814 IL10RB as a key regulator of COVID-19 host susceptibility and severity

PrgmNr 2815 A genome-wide association study identifies two novel loci for respiratory infection with Pseudomonas aeruginosa in cystic fibrosis

PrgmNr 2816 Aberrant levels of CXCL16 in severe COVID19 patients

PrgmNr 2817 Adding polygenic risk scores drastically improves the diagnostic accuracy of conventional lab tests for systemic lupus erythematosus

PrgmNr 2818 Application of a novel instrument-free and microfluidics-free single-cell analysis technology (PIPseq) that is well suited for viral applications in resource constrained laboratories

PrgmNr 2819 Assessing the potential correlation of polymorphisms in the IL6R gene with relative IL-6 elevation in severely ill COVID-19 patients

PrgmNr 2820 Comparative genetic architecture of the inflammatory bowel disease across East Asian and European populations

PrgmNr 2821 COVID-19 Dynamic pQTLs: Genetic Determinants of Proteins that are Unique to SARS-CoV-2 Infection

PrgmNr 2822 Data-driven analysis exploring age and sex effects on gene expression for dengue hemorrhagic fever

PrgmNr 2823 Downregulation of exhausted cytotoxic T cells in gene expression networks of multisystem inflammatory syndrome in children

PrgmNr 2824 Fine-mapping of multiple sclerosis susceptibility loci in a multi-ethnic population

PrgmNr 2825 Genetic and expression changes associated with recurrence in the perioperative ileal resection period

PrgmNr 2826 Genome-wide analysis of schizophrenia and multiple sclerosis reveals shared genetic loci outside the major histocompatibility complex

PrgmNr 2827 Genome-wide association study of juvenile idiopathic arthritis identifies novel risk loci

PrgmNr 2828 Genomic surveillance of SARS-CoV-2 variants for the center of disease control using full viral genome sequencing in 12,000 confirmed patient saliva specimens

PrgmNr 2829 Identification of miR-eQTLs from small RNA Seq in Children with Asthma

PrgmNr 2830 The Genetics of The Antibody Response to Human Adenovirus infection in Healthy Individuals

PrgmNr 2831 Comprehensive genetic screening for variants associated with spermatogenic failure

PrgmNr 2833 Glucocorticoid Regulated KCNA5 Mediates Cell Cycle in Fetal Membranes

PrgmNr 2834 Reprograming human fibroblasts into Sertoli cells: a tool for personalized medicine

PrgmNr 2835 The identification of disease complication risk via disease-disease network: an application to obstetric disorders in the UK Biobank

PrgmNr 2836 Genome-wide association study and gene-based analysis of participants with hemophilia A and inhibitors in the My Life Our Future study

PrgmNr 2837 Integrative analyses of germline variation and somatic mutations identify germline-somatic interactions in myelofibrosis susceptibility

PrgmNr 2838 MAPT Structural Haplotypes in Chronic Traumatic. Encephalopathy

PrgmNr 2839 Autism spectrum disorder phenotype subset in attention deficit hyperactivity disorder: An analysis of first-tier test results and phenotype spectrum and severity

PrgmNr 2840 Brain trans-pQTL analysis of GWAS loci reveals novel link between psychiatric disorders and neurodegenerative diseases

PrgmNr 2841 Copy number variation identification and association study on 3,800 Alzheimer's disease whole genome sequencing data from the Alzheimer's Disease Sequencing Project (ADSP)

PrgmNr 2842 Exome Sequencing Identifies Variants on the X-Chromosome Associated with Alzheimer’s Disease in the Alzheimer’s Disease Sequencing Project

PrgmNr 2843 Genetic Variant, rs2304672, in 5’ UTR region in hPER2 gene and its association to Parasomnias and other sleep disorders in Puerto Rico

PrgmNr 2844 Predicting Mortality Among Ischemic Stroke Patients Using Polygenic Risk Scores Derived From the Disease-Related Biological Pathways

PrgmNr 2845 Replicating genetic associations in Alzheimer's Disease using data from Eisai's BAN2401 ABBA 201 clinical trial

PrgmNr 2846 Severity of dementia ratings in Hispanic Latinos as a function of country of origin and APOE

PrgmNr 2847 SnRNA-seq probing of the differential vulnerability of motor neurons in ALS

PrgmNr 2848 The influence of APOE e4 and global ancestry on self-reported anxiety and depression in Hispanic, Non-Hispanic Whites, and African Americans

PrgmNr 2849 Two Genetic Variants Explain the Association of European Ancestry with Multiple Sclerosis Risk in African-Americans

PrgmNr 2850 Two-trait integrative analysis of de novo mutations in autism spectrum disorder and developmental disorders identifies shared CNS processes and epigenomic functions

PrgmNr 2852 Assortative mating drives pathogenicity of variably expressive CNVs

PrgmNr 2853 Clinical value of adding copy number, tandem repeat and antipsychotics metabolizer status analysis to diagnostic sequencing of patients with psychotic disorders

PrgmNr 2854 Contribution of copy number variants to schizophrenia in East Asian populations

PrgmNr 2855 Dissecting Biological Pathways of Psychopathology using Cognitive Genomics

PrgmNr 2856 Enrichment of the VNTR PER3 variant in DSWPD patients: A large whole genome sequencing analysis

PrgmNr 2857 Extracellular vesicles: A window into the etiology of major depressive disorder

PrgmNr 2858 Genome-wide genetic susceptibility to pain and depression predicts suicidal thoughts and attempts inschool-agedchildren

PrgmNr 2859 Identifying major depressive disorder subtypes using polygenic risk scores

PrgmNr 2861 Pathways Implicated in Risk for Suicide Attempts in the Million Veterans Program

PrgmNr 2862 Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia, Bipolar Disorder, and Major Depression in 660,000 US Veterans

PrgmNr 2863 Perspectives on the COVID-19 pandemic: Report from individuals with copy number variations and their caregivers

PrgmNr 2864 Suicide attempts among European ancestry veterans in the Million Veteran Program exhibits correlation with multiple psychiatric conditions

PrgmNr 2865 Whole exome sequencing meta-analysis of depression suggests subtle role for functional variants in known GWAS loci

PrgmNr 2866 Whole-exome sequencing in the UK Biobank reveals risk gene SLC2A1 and biological insights for major depressive disorder

PrgmNr 2867 A genetic epidemiologic study of myalgic encephalomyelitis (ME)/chronic fatigue syndrome (CFS) identifies links with autoimmune disease and cancer

PrgmNr 2868 A multi-model data approach to improve statistics for Metabolite-Variant-Disease (MVD) relationship in SOL dataset

PrgmNr 2869 A reliable algorithm to reconstruct modifier haplotypes in family-based studies of complex disorders using variable ages-at-onset for primary open-angle glaucoma as a disease model

PrgmNr 2870 Association of Clonal Hematopoiesis with Chronic Obstructive Pulmonary Disease

PrgmNr 2871 Exome sequencing and primary care data from the UK Biobank reveal an opportunity for an RNAi therapeutic targeting XDH to treat gout

PrgmNr 2872 Gene-environment interactions determine human milk oligosaccharide composition in lactating mothers

PrgmNr 2873 Generalizing Kernel Association Tests with a Protein Structure Based Kernel

PrgmNr 2874 Genetic associations of protein-coding variants in human disease

PrgmNr 2875 Genetically increased circulating FUT3 level is associated with reduced risk of Idiopathic Pulmonary Fibrosis: a Mendelian Randomization Study

PrgmNr 2876 Genetics of male pattern baldness in sub-Saharan Africa

PrgmNr 2877 Genome-wide association and comparative gene set enrichment analysis identifies both similarities and differences between skeletal growth phenotypes

PrgmNr 2878 Genome-wide inter-chromosomal epistatic associations identified across complex diseases in the ~300,000 participants from eMERGE and UK Biobank

PrgmNr 2880 Human milk oligosaccharides alter the milk microbiota in lactating mothers and impact the gut microbiota of breastfed infants with asthma

PrgmNr 2881 Identifying synthetic interactions between synonymous mutations in the UK Biobank WES data using REVEAL: Biobank

PrgmNr 2882 Incident disease associations with mosaic chromosomal alterations on autosomes, X and Y chromosomes: insights from a phenome-wide association study in the UK Biobank

PrgmNr 2883 Integration of a MUC5B Promoter Variant and a Polygenic Risk Score for Idiopathic Pulmonary Fibrosis and Interstitial Lung Abnormalities

PrgmNr 2884 Investigating the transferability of European height associated loci in Africans

PrgmNr 2885 Large-scale exome sequencing identifies 8 genes impacting adult cognitive function

PrgmNr 2886 Neuroimaging-guided functional variant multi-ancestry PheWAS using electronic health records from UK Biobank

PrgmNr 2887 Organelles and aging: a human genetics approach

PrgmNr 2888 Pakistan Genomic Resource: The world’s largest biobank of human knockouts

PrgmNr 2890 Polygenic gene-environment interactions: from genetic architecture to pharmacogenomics

PrgmNr 2891 Prioritizing Research Variants in the NIH Undiagnosed Diseases Program

PrgmNr 2892 Projecting genetic associations and drug transcriptional profiles through gene expression patterns reveal disease etiology and potential mechanisms for therapeutic strategies

PrgmNr 2893 Quantifying factors that affect polygenic risk score performance in the eMERGE dataset

PrgmNr 2894 Rare variant GWAS of African American and Hispanic patients using WGS data from the TOPMed program on the NHLBI BioData Catalyst

PrgmNr 2895 Rare, functional variants amongst 180,256 exomes from the UK Biobank influence mitochondrial copy number

PrgmNr 2896 Relationship between autozygosity and complex traits in the Amish

PrgmNr 2897 Single cell enrichment of expression and splicing QTL target genes mapped to GWAS loci identifies causal genes and pathogenic cell types for glaucoma and intraocular pressure

PrgmNr 2898 Top-LD: a tool to explore linkage disequilibrium using TOPMed whole genome sequence data

PrgmNr 2899 Trans-ethnic Transcriptome-wide Association Study and fine-mapping analysis in 3.4 million individuals shed lights on the genetic architecture of alcohol and smoking addiction

PrgmNr 2900 Whole exome sequencing analysis identified five novel genes associated with Osteoarthritis

PrgmNr 2901 A comprehensive functional screen detects multiple polymorphic RET enhancers affecting Hirschsprung disease risk

PrgmNr 2902 Association of DNA Methylation with Mixed Substance Use in an HIV-positive Cohort

PrgmNr 2903 Cis-regulatory hubs constitute a powerful model to understand the impact of 3d organization in schizophrenia

PrgmNr 2904 Developmental regulation of neuronal gene expression by Elongator complex protein 1 dosage

PrgmNr 2905 Differential methylation signatures in atypical parkinsonism syndromes

PrgmNr 2906 Dissecting multiple-signal GWAS loci and their regulatory roles in Alzheimer’s Disease

PrgmNr 2907 DNA methylation signature of ASXL1 variants causing Bohring-Opitz syndrome (BOS)

PrgmNr 2908 Epigenome-wide association study of circulating IgE levels identifies novel targets for asthma

PrgmNr 2909 Genetic effects on brain traits impact cell-type specific alternative splicing during human neurogenesis

PrgmNr 2910 Leveraging the Mendelian Disorders of the Epigenetic Machinery to Systematically Map Functional Epigenetic Variation

PrgmNr 2911 MeQTL Mapping for Present Cocaine Use and Persistent Cocaine Use in a Veteran Population

PrgmNr 2912 Meta-analysis of gene expression data in Alzheimer's disease identifies sex-differential effects in multiple brain tissues

PrgmNr 2913 Population-level variation of enhancer expression identifies novel disease mechanisms in the human brain

PrgmNr 2914 Power-improved meta-QTL analysis reveals the complex regulation of molecular QTL associated with brain diseases

PrgmNr 2915 Sex differences in the human brain transcriptome of cases with schizophrenia

PrgmNr 2916 Sex-specific profiles of m6A RNA methylation in the brain of individuals with major depressive disorder

PrgmNr 2917 Single Nuclei Sequencing of Human Putamen Oligodendrocytes Reveals Altered Heterogeneity and Disease-Associated Changes in Parkinsons Disease and Multiple System Atrophy

PrgmNr 2918 TET1 mediated modulation of Alzheimer’s disease

PrgmNr 2919 The Musculoskeletal 3D Epigenome Atlas

PrgmNr 2920 Unraveling the role of cell type specific noncoding variations in craniofacial disease

PrgmNr 2921 Widespread shifts in muscle gene expression associated with reduced gait speed in a non-human primate model of aging

PrgmNr 2922 Accurate identification of circRNA landscape and dynamic regulation during early neuronal differentiation

PrgmNr 2923 Alcohol Use Disorder is associated with DNA methylation-based shortening of telomere length and regulated by TESPA1: implications for aging

PrgmNr 2924 Chromatin accessibility and gene expression during adipocyte differentiation identify context-dependent effects at cardiometabolic GWAS loci

PrgmNr 2925 Deconvolution of genetic variation using high-quality cis-regulatory elements map of kidney cells

PrgmNr 2926 Engineering a synthetic humanized RET mouse to model Hirschsprung Disease

PrgmNr 2928 Functional characterization of regulatory elements involved in mouse CD4+ T cell differentiation

PrgmNr 2929 Human microglia regulome analysis defines the inherited risk loci in Alzheimer's Disease

PrgmNr 2930 Identification of bidirectional regulatory region between FADS1 andFADS2

PrgmNr 2931 Incorporating local ancestry improves the identification of ancestry-associated methylation signatures and meQTLs in the African American population

PrgmNr 2932 Sex differences in the intergenerational link between maternal and neonatal whole blood DNA methylation: An analysis in the Boston Birth Cohort

PrgmNr 2933 Studying Regulatory Variation In Founder Populations To Identify Functional Rare Variants

PrgmNr 2934 Understanding the interplay of pancreatic cancer GWAS risk loci and cellular stress

PrgmNr 2935 Widespread sex differences in placental DNA methylation

PrgmNr 2936 Wnt sensitive regulatory elements in neural progenitors harbor neuropsychiatric disorder and brain structure heritability

PrgmNr 2937 Accurate identification of circRNA landscape and microRNA network in human oligodendroglia differentiation

PrgmNr 2938 Integrative multi-omics analysis to predict gene regulatory networks from genetic risk variants to phenotypes of Alzheimer's disease and COVID-19

PrgmNr 2939 Learning interpretable cellular and gene signature embeddings from single-cell transcriptomic data

PrgmNr 2940 ModTools: a computational toolbox for rapid detection of DNA modifications and replications using Nanopore sequencing

PrgmNr 2941 De novo mucins in mammals: How parallel evolution of exonic repeats leads to new function

PrgmNr 2942 A spatially aware likelihood test to detect sweeps from haplotype distributions

PrgmNr 2943 A systematic evaluation of phasing algorithms for downstream analysis in ancestry testing

PrgmNr 2944 African ancestry polygenic risk scores improve Alzheimer disease risk prediction in individuals of African Ancestry

PrgmNr 2945 Dissection of ethnic and socio-cultural substructure of the South Asian Indian population using genetic data

PrgmNr 2947 Frequency enrichment of functional coding variants in a French Canadian founder population and its implication for inflammatory bowel diseases

PrgmNr 2948 Gene duplication analysis reveals genome evolution and adaptation of Taenia species

PrgmNr 2949 Genomic patterns of natural selection and introgression at the alcohol dehydrogenase gene region are associated with agriculture in ethnically diverse Africans

PrgmNr 2951 Mutational biases in the SARS-CoV-2 virus genome

PrgmNr 2952 Novel unannotated protein-coding genes are expressed in multiple tissues

PrgmNr 2953 Overcoming constraints on the detection of recessive selection in human genes from population frequency data

PrgmNr 2954 PLIGHT: A tool to assess privacy risk by inferring identifying characteristics from sparse, noisy genotypes

PrgmNr 2955 RaPID-Query: Towards a Fast and Accurate Real-Time Genealogical Search

PrgmNr 2956 A Decision Science Approach to Understanding Motives that Underly Interest in Genomic Sequencing Results

PrgmNr 2957 Addressing Barriers to Access in an Adult Kidney Genetics Clinic

PrgmNr 2958 Addressing barriers to patient data sharing: Exploring the effects of employing electronic consent to obtain genetic testing records through ClinGen's Patient Data Sharing Program

PrgmNr 2959 Benefits, harms and costs of newborn genetic screening for hypertrophic cardiomyopathy: estimates from the PreEMPT Model

PrgmNr 2960 Caregivers' understanding of 22q11 Deletion Syndrome and their communication with their affected adult children about the syndrome

PrgmNr 2961 Clinical benefits, cost-effectiveness and value of further research for population-based newborn screening for Li-Fraumeni syndrome

PrgmNr 2962 Communication about risk, testing, and test results with and through family members: A survey of families with Huntington's disease

PrgmNr 2963 Cost-efficient risk assessment for COVID-19 at work and play using machine learning

PrgmNr 2964 Deciding to pursue pediatric whole genome sequencing: Exploring value from parents' perspectives

PrgmNr 2965 Diagnostic sequencing to support genetically stratified medicine in a tertiary care setting

PrgmNr 2966 Digital tools for delivering genomic services: A systematic review

PrgmNr 2967 Emerging public policy and perspectives on biometrics (including DNA and facial imaging) in the US

PrgmNr 2968 Ethical issues in pediatric gene therapy research

PrgmNr 2969 Genes to Mental Health Network Report: Stakeholder perspectives on research participation

PrgmNr 2970 Hearing in Generation Genome: Comprehensive Newborn Hearing Screening through SEQaBOO (SEQuencing a Baby for an Optimal Outcome)

PrgmNr 2972 NHGRI's Inter-Society Coordinating Committee for Practitioner Education in Genomics: Multi-disciplinary genomics resource development

PrgmNr 2973 Polygenic risk scores change primary care providers' preventive care of racially diverse patients: Results of a national survey with randomized case scenarios

PrgmNr 2974 Population carrier frequency estimates may improve genetic counselling for rare autosomal recessive syndromes due to biallelic variants in cancer risk genes

PrgmNr 2975 Preparing primary care providers to utilize direct-to-consumer genetic test results in the creation of an individualized healthcare plan

PrgmNr 2976 Prevalence and Correlates of Parental Decisions about Direct-to-Consumer Genetic Testing for Adult-Onset Inherited Cancer Syndromes in their Children

PrgmNr 2977 Public knowledge, perceptions, and attitudes towards direct-to-consumer genetic testing: A national web-based survey

PrgmNr 2978 Returning Polygenic Risk Score-Based Results to Parents - Parents' Understanding of and Response to Reported Risk

PrgmNr 2979 Secondary Findings in Minors Case Series: Recommendations for Clinical Care and Research

PrgmNr 2980 Strategies to address the health and social care needs of rare disease patients: A comparative analysis of two Canadian provinces

PrgmNr 2981 The complexity of diagnosing rare disease: An organising framework for health services and health economics research based on real world evidence

PrgmNr 2982 The cost of good health: Poverty association with differential gene expression

PrgmNr 2983 The Effect of Genetic Testing as Part of Personalized Lifestyle and Habit Change Coaching

PrgmNr 2984 The physical costs and psychosocial benefits associated with moving to a White neighborhood for African American male children

PrgmNr 2985 The White Ceiling: Qualitative Study of Genetic Counseling Students' Perceptions about Cultural Competence Training

PrgmNr 2986 Views on the training needs of genetic assistants vary widely

PrgmNr 2987 Yield of genomic disease given the identification of a secondary finding

PrgmNr 2988 A 3-part Phase 2 study of HST5040, an investigational oral therapy that reduces toxic coenzyme A metabolites in methylmalonic and propionic acidemias (clinicaltrials.gov NCT04732429)

PrgmNr 2989 Characterization of CRISPR gene editing in the brain of ataxia mouse models with targeted PCR-free Nanopore sequencing

PrgmNr 2990 CRISPR targeting of the dominantCochp.A449T mutation in a patient-derived skin fibroblast cell line and in a new humanized mouse model

PrgmNr 2991 Haplotyping SNPs for allele-specific gene editing of the mutant huntingtin allele using long-read sequencing

PrgmNr 2992 Human genetic diversity modifies therapeutic gene editing off-target potential

PrgmNr 2993 Long-term, sustained efficacy and safety from a phase 1/2 clinical trial of an AAV8-mediated liver-directed gene therapy in adults with glycogen storage disease type Ia

PrgmNr 2994 NMD-dependent approach restores CFTR function in primary nasal cells harboring nonsense variants

PrgmNr 2995 Patient-focused drug development for a single intravenous dose of HMI-203 gene therapy in adult mucopolysaccharidosis (MPS) II, or Hunter syndrome, patients

PrgmNr 2996 Somatic CAG expansion in Huntington's disease is dependent on the MLH3 endonuclease domain, which can be excluded via splice redirection

PrgmNr 2997 Sorbitol reduction in SORD deficient fibroblasts and iPSC-derived motor neurons using an antisense oligonucleotide (ASO) treatment strategy

PrgmNr 2998 Towards precision genome editing in an inborn error of polyamine metabolism

PrgmNr 2999 A deeply sequenced public resource of diverse human genomes

PrgmNr 3000 A web-based application interface for UK Biobank phenotype exploration

PrgmNr 3001 An Extensible Prototype for MultiOmic Clinical Reporting

PrgmNr 3002 Characterizing repeat expansion variation in the Undiagnosed Disease Network cohort

PrgmNr 3003 Development of a comprehensive, locus specific, database for ENPP1 Deficiency (generalized arterial calcification of infancy/autosomal recessive hypophosphatemic rickets (GACI/ARHR2) to clarify the clinical relevance of variant data

PrgmNr 3004 Examination of clinical features among 200,632 UK Biobank participants harboring deleterious germline variants in NF1 and SPRED1

PrgmNr 3005 Improving RNAseq Splice Junction Prioritization in Rare Disease Analysis

PrgmNr 3006 Improving the imputation quality of French-Canadian genomic data

PrgmNr 3007 Making Discoveries with Kids First Variant Database

PrgmNr 3008 NCBI ALFA Release 2for900 Million Variantsand Allele Frequencyfrom 200KdbGaPSubjects

PrgmNr 3009 NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS): 2021 Update

PrgmNr 3010 Phenome-wide association analysis of rare and common variation in 455,000 UK Biobank exome sequences accessible through public data portal

PrgmNr 3011 Real-world Evidence for Human Phenotype Ontology Concept Prevalence and its Association with Genetic Disorders from Electronic Health Records

PrgmNr 3012 Tools and Resources to Improve Understanding of Rare Genetic Conditions in Newborn Screening

PrgmNr 3013 Toward a gold-standard set for SNP to gene mapping

PrgmNr 3014 An individual with craniosynostosis, hypertelorism, and imperforate anus with normal cognition caused by a 7.4 Mb duplication on chromosome 6

PrgmNr 3015 Development of a Genetic Diagnostic Algorithm for Individuals with Split Hand Foot Malformation

PrgmNr 3016 Missense variants affecting the actin-binding domains of PLS3 cause X-linked congenital diaphragmatic hernia and body wall defects

PrgmNr 3017 Multiple independent genetic syndromes in a single patient

PrgmNr 3018 Seckel-6 Syndrome: A new severe phenotype

PrgmNr 3019 The Expanding Phenotype of WAGR Syndrome: Reconceptualization from a Syndrome to a Spectrum Disorder

PrgmNr 3020 ADAM6 (106329183_106736911)x3 gene variant in a Charcot Marie-Tooth patient with upper extremity involvement

PrgmNr 3021 SETBP1 involvement in a neurodevelopmental disorder: A report from a reanalysis of a negative clinical exom

PrgmNr 3022 TBCK-related encephalopathy: Understanding the role of autophagy in the pathology of the disease

PrgmNr 3023 A cross-disorder and evidence-based tiered ranking of candidate genes for neuropsychiatric disorders

PrgmNr 3024 A late-onset seizure phenotype in a mouse model of CHD8 haploinsufficiency

PrgmNr 3026 A screening pipeline to identify functional impacts of novel KCNT1 variants in pediatric epilepsies

PrgmNr 3027 Autism spectrum disorder: Magel2 gene variants

PrgmNr 3028 Biallelic inframe deletion of human SOX4 is associated with developmental delay, hypotonia and intellectual disability

PrgmNr 3029 Dysregulation of Canonical and Alternative Replication Protein A Complexes in Huntington Disease and Spinocerebellar ataxia Type 1 brains is associated with CAG instability and phenotype

PrgmNr 3030 Evaluating digenic inheritance of bi-allelic PGAP2 and PGAP3 variants in a case of Mabry syndrome

PrgmNr 3031 Fibroblasts and lymphocytes are superior options to whole blood and other peripheral tissues for Mendelian neurological disease diagnosis through RNA-seq

PrgmNr 3032 Free Sialic Acid Storage Disease (FSASD) due to defectiveSLC17A5: A status update

PrgmNr 3033 Genetic evaluation of dementia with Lewy bodies implicates distinct disease subgroups

PrgmNr 3034 Genome wide meta-analysis of suicide behaviors

PrgmNr 3035 Heterozygous Loss-of-Function Variants in TBCK Cause a Mild Neurologic Syndrome in Humans and Mice

PrgmNr 3036 Identification of a Novel CNTNAP2Variant in a Consanguineous Pakistani Family with Epilepsy

PrgmNr 3037 Impaired SNAPC4function leads to global reduction of canonical splicing events and is associated with a disorder characterized by progressive spasticity, developmental delay, and speech dysarthria

PrgmNr 3038 Neurofilament light chain in cerebrospinal fluid as a potential novel biomarker in evaluating both clinical severity and therapeutic response in Niemann-Pick Disease, type C1

PrgmNr 3039 Pre-clinical studies in induced Pluripotent Stem Cell (iPSC) lines with SORD mutations linked to a recessive neuropathy

PrgmNr 3040 Truncating variants in the SHANK1 gene are associated with a spectrum of neurodevelopmental disorders

PrgmNr 3041 UsingDrosophila melanogasterto identify loci modifying Coffin-Siris syndrome mutations

PrgmNr 3042 Biochemical consequence of deficient glycan extension in ALG3-CDG

PrgmNr 3043 Bone cell functions in PPIB knock-out mouse model for type IX osteogenesis imperfecta are distinct from classical dominant OI

PrgmNr 3044 Distribution and frequency of the variant rs10738445 of BNC2 gene associated to Idiopathic Scoliosis in Puerto Rico

PrgmNr 3045 GPC6 haploinsufficiency in patients with skeletal anomalies
PrgmNr 3046 Molecular test of Paget's disease of bone in families not linked to SQSTM1 gene mutations
PrgmNr 3047 Neonatal presentation of Methylmalonic Acidemia (MMA) with only a single heterozygous mutation in MUT
PrgmNr 3049 Prevalence and Type of Gastrointestinal Symptoms in the Adult Skeletal Dysplasia Population
PrgmNr 3050 RNA-seq identifies novel regulatory variants underlying Glycogen Storage Diseases
PrgmNr 3051 Spectrum of Germline and Somatic Mitochondrial DNA Variants in Tuberous Sclerosis Complex
PrgmNr 3052 7 day-old boy presenting with anuria and electrolyte imbalance: Novel, disease-associated variant identified in SLC5A1 gene
PrgmNr 3053 Ascertainment context is crucial: The penetrance of age-dependent monogenic disease variants depends on ascertainment context
PrgmNr 3054 Challenges in Classifying Clinically Relevant Germline RTEL1 Variants
PrgmNr 3055 Colorblindness gene implicated for myopia in highly aggregated Pennsylvania Amish families
PrgmNr 3056 Comparative analysis of the contribution of copy number and single nucleotide variants to the pathogenesis of idiopathic hypogonadotropic hypogonadism
PrgmNr 3057 Density Volumetric Analysis of Cystic Lung Disease in Proteus Syndrome
PrgmNr 3058 Determining the role of Frizzled pathway candidates genes in causing the rare inherited blinding disorder FEVR
PrgmNr 3059 Diagnostic yield of CNV analysis in primary immunodeficiencies
PrgmNr 3060 Exome sequencing findings in patients with bronchiectasis within an immune system disorder cohort
PrgmNr 3061 Genomic Analysis of Profound Childhood Hearing Loss in the Yoruba Population of Nigeria
PrgmNr 3062 Intronic mutations in Puerto Rican (Hispanic) children with rare genetic diseases: Understanding genotype and phenotype correlations at the individual level
PrgmNr 3063 Mosaicism in a cohort of patients with immune disorders
PrgmNr 3064 The complexities of germline PARN variants in telomere biology disorders and the population
PrgmNr 3065 Update to the NIAID Genomic Research Integration System Genomic Analysis Module - Lessons Learned
PrgmNr 3067 3C minus 2C: X-linked Ritscher-Schinzel/ Cranio-Cerebello-Cardiac syndrome with no cardiac or cerebellar malformation due to missense variant inCCDC22
PrgmNr 3068 45,X/46,XY Mosaicism: Retrospective Study of 77 Patients
PrgmNr 3069 Case report: a boy with imperforate anus and cleft lip
PrgmNr 3070 Diagnoses and discoveries from re-analysis of clinical exome sequencing data: The Care4Rare Canada experience
PrgmNr 3071 Exome CNV calling and analysis in a large cohort of families with undiagnosed rare genetic disease
PrgmNr 3072 Phenotypic spectrum of Down syndrome in 4,209 patients by profiling electronic health records
PrgmNr 3073 A partial Xq duplication in a newborn girl: A Case Report
PrgmNr 3074 A rare case of multiple mosaic marker chromosomes in a neonate with congenital diaphragmatic hernia and ventricular septal defect
PrgmNr 3075 Assessment of Alzheimer Disease Related Plasma Biomarker Phosphorylated Tau 181 in Individuals of Diverse Ancestral Backgrounds
PrgmNr 3076 Clinical utility of long read sequencing in resolving inconclusive diagnoses due to gene conversions and unknown phase
PrgmNr 3077 Comprehensive genetic testing of sensorineural hearing loss in children: understanding phenotypic factors that influence the diagnostic yield

PrgmNr 3078 Genome-wide Sequencing Ontario (GSO): An implementation pilot to improve rare disease diagnostics

PrgmNr 3079 Going above and beyond in discovering STRC pathogenic variants in non-syndromic hearing loss: it's just the tip of the iceberg

PrgmNr 3080 Molecular diagnosis as an incidental finding from preconception carrier screening

PrgmNr 3081 Non-invasive diagnostic testing for uterine fibroids: A cell-free DNA approach

PrgmNr 3082 Optical Genome Mapping for Constitutional Postnatal SV, CNV, and Repeat Array Sizing: A Multisite Clinical Validation Study

PrgmNr 3083 To report or not to report: The power of curating gene-disease relationships with limited evidence for rare disease

PrgmNr 3084 Kmt5b is highly expressed in the developing brain and may regulate other known autism risk genes and processes

PrgmNr 3085 TRAPPC10 mutation is associated with a microcephalic TRAPPopathy disorder in humans and mice

PrgmNr 3086 Advanced Diagnostics and Genotype-Phenotype Resolution using Functional Genomics in >500 Neuromuscular and Neurological Disorder Patients

PrgmNr 3087 Cell culture GWAS identifies common genetic variants that influence lithium induced neural progenitor proliferation

PrgmNr 3088 Characterization of molecular effects of a novel GNAQ mutation identified in Sturge-Weber Syndrome

PrgmNr 3089 Clinical and biochemical characteristics of MOGS-CDG, a rare congenital disorder of glycosylation

PrgmNr 3090 CRISPR engineering of a dystonia-specific allelic series of THAP1 mutations reveals transcriptional profiles associated with myelination and neurodevelopment

PrgmNr 3091 Distribution and allele frequency of theGln3060Xvariant present in theASPMgene associated to microcephaly and related disorders in Puerto Rico

PrgmNr 3092 Enhancement of CHRNA5 mRNA expression by upstream polymorphisms associated with preserved cognition and lessened neuropathology in an aged human population

PrgmNr 3093 Functional characterization of haplotype surrounding TOMM40-523' repeat to assess differential risk effects on European ancestry APOEε3 haplotypes

PrgmNr 3094 Identification of candidate genes for alcohol use disorder using RNA-seq data

PrgmNr 3095 Influence of APOE e4in neuropathologic lesions of Alzheimer disease and related dementias

PrgmNr 3096 Molecular mechanisms of autism-associated variants in KCNQ3 channel

PrgmNr 3097 Multi-region single-cell dissection of brain vasculature changes in Alzheimer's Disease

PrgmNr 3098 Patterns of gene expression variation in large families carrying developmental delay-associated 16p12.1 deletion

PrgmNr 3099 Plasma membrane injury may trigger DUX4 expression in FSHD

PrgmNr 3100 Re-analysis of regulatory non-coding promoter/enhancer regions of known congenital myasthenic syndrome genes to identify disease-causing variants

PrgmNr 3101 Sex bias in neurodevelopmental defects due to differential effects of gene dosage and genetic interactions

PrgmNr 3102 Spontaneous DNA damage response, myotonic dystrophy type 1, and CTG tract size

PrgmNr 3103 Variant analysis in Gaucher sibling pairs discordant for parkinsonism to identify secondary genetic risk factors for GBA1-associated Parkinson disease

PrgmNr 3104 X-linked recessive mutations in PDZD4 are likely associated with neurodevelopment delay and autism spectrum disorder

PrgmNr 3105 Cross-disorder study of neuropsychiatric CNVs and polygenic scores in a large health system-based population

PrgmNr 3106 Infant with PMM2-CDG(CDG-1a) masquerading as infectious pericarditis

PrgmNr 3107 A scalable bioinformatic approach to recover variation in camouflaged gene regions in whole genome sequencing data from the UK Biobank

PrgmNr 3108 Accurate assignment of disease liability to genetic variants using only population data

PrgmNr 3109 Alternative polyadenylation quantitative trait loci (APA-QTLs) in chronic obstructive pulmonary disease

PrgmNr 3110 Analysis of splice site variation using PEX-DETEX indicates an enrichment of poison exons in individuals with developmental brain disorders

PrgmNr 3111 CNV-Disease Curation and Role of Metabotropic Glutamate Receptors Optimized for Neurodevelopmental Disorders with ParseCNV2

PrgmNr 3112 Evaluation of splicing in F9 and ADAMTS13 Exonic Synonymous Variants

PrgmNr 3113 Exploring the genotypic and phenotypic significance of KCNJ1 variants in UK Biobank data using REVEAL: Biobank

PrgmNr 3114 Genetic control of mRNA splicing is affected by purifying selection and may be linked to incomplete penetrance

PrgmNr 3115 Identification of discriminative gene-level and protein-level features associated with gain-of-function and loss-of-function mutations

PrgmNr 3116 Identification of polyadenylation signals relevant to Mendelian disease variant interpretation

PrgmNr 3117 Integrated analysis of co-expression and exome sequencing to prioritize susceptibility genes for cutaneous melanoma

PrgmNr 3118 Loss-of-function tolerance in human disease genes

PrgmNr 3119 LUSTR: A powerful and user-friendly tool to call germline and somatic short tandem repeat variants from short read next-generation sequencing data

PrgmNr 3121 Rare variants affecting multi-omic measurements of gene regulation implicated in EKG traits

PrgmNr 3122 Regional missense mutational tolerance in 316,810 exomes

PrgmNr 3123 Transcription factor analysis across 3,604 GWAS uncovers novel pathway-specific associations

PrgmNr 3124 Transcriptome-wide association studies and fine-mapping at cell-type resolution

PrgmNr 3126 Why do males die? Uncovering the genetic connections causing COVID-19 disease severity

PrgmNr 3127 CRISPR activation at a trait-associated SNP reveals trans effects on gene expression

PrgmNr 3128 Insights into Molecular Pathways Associated with Juvenile Myositis through Single-Cell Transcriptome Profiling

PrgmNr 3129 Proteomics study on bone biopsy sample

PrgmNr 3130 Single-cell and spatial transcriptomics of reovirus induced myocarditis

PrgmNr 3132 ACME: An Affinity-based Cas9 Mediated Enrichment method for targeted nanopore sequencing

PrgmNr 3134 Classification of Super populations based on multidimensional scaling analysis of SNP array data

PrgmNr 3135 Genomic assessment of single bacteria with Primary Template-directed Amplification for deconvolution of microbial diversity

PrgmNr 3136 Improved first-strand cDNA synthesis kits incorporating qScript Ultra, a novel, engineered, thermostable reverse transcriptase enabling full-length synthesis of cDNA over 20 kb in 10 minutes

PrgmNr 3137 Learning single-cell embeddings for bulk tissue cellular deconvolution

PrgmNr 3138 Long read sequencing reveals extensive structural variation in diverse mouse genomes

PrgmNr 3139 MERSCOPETM reveals the transcriptional organization of the mouse brain

PrgmNr 3140 Normalizing UDI Library Construction for Sensitive Genomic Applications

PrgmNr 3141 Outcome of 1500 matches through the Matchmaker Exchange for rare disease gene discovery: The 2-year experience of Care4Rare Canada

PrgmNr 3142 sparQ RNA-Seq Kit: Improved RNA sequencing with 4.5 hr library prep using a novel, integrated ribo-globin depletion technology

PrgmNr 3143 TRPC6 gain-of-function in doxorubicin-induced heart failure

PrgmNr 3144 Paradigm Shift from Disease PRS to PGx PRS for Drug Response Prediction using PRS-PGx Methods
PrgmNr 3145 Patterns ofCYP2D6Phenotypes and the impact on precision medicine: New resource from Medical Genetics Summaries
PrgmNr 3146 Pharmacogenomic polygenic risk score for clopidogrel responsiveness among Caribbean Hispanics
PrgmNr 3147 Results of a Real-world implementation of a pharmacogenetic-empowered comprehensive medication management (CMM) program
PrgmNr 3148 Haplotype-aware inference of human chromosome abnormalities
PrgmNr 3149 Human milk oligosaccharides protect lung health in breastfed infants of the CHILD Study
PrgmNr 3150 Impact of maternal nutrition on transcriptome changes over time during fetal liver development
PrgmNr 3151 Increased number of de novo mutations in craniofacial regulatory elements in trios with orofacial clefts
PrgmNr 3153 Multi omics integration in placenta identifies candidate functional genes for birthweight
PrgmNr 3154 The Effect of Variable Hormonal Replacement Therapies (HRT) on Neurocognition in School-aged Males with 47,XXY
PrgmNr 3155 The Longitudinal Effect of Early Hormonal Treatment (EHT) in Males with 47,XXY
PrgmNr 3156 Trans-ethnic meta-analysis of genome-wide association studies identifies novel loci for gestational duration
PrgmNr 3157 Whole mitochondrial genome analysis of a single trophectoderm biopsy - heteroplasmy dynamics and selection in human embryos
PrgmNr 3158 TESLA: super-resolution Tumor Edge Structure and Lymphocyte multi-level Annotation from spatial transcriptoms
PrgmNr 3159 A fast and robust Bayesian nonparametric method for prediction of complex traits using summary statistics
PrgmNr 3160 A fresh look at the role of Hardy-Weinberg disequilibrium in association testing
PrgmNr 3161 A novel regression-based method for X-chromosome-inclusive Hardy-Weinberg equilibrium test
PrgmNr 3162 A sparse high-dimensional generalized varying coefficient model for identifying genetic variants associated with regional methylation
PrgmNr 3163 AFA Computationally efficient Ancestral Frequency estimation in Admixed populations: the Hispanic Community Health Study/Study of Latinos
PrgmNr 3164 Applying Recurrent Weighted Replanting to detect gene-gene interaction in case-parent trios
PrgmNr 3165 Approaches for detection of epistatic interactions of causal variants in genome-wide data: Comparison of Recurrent Weighted Replanting with other machine learning methods
PrgmNr 3166 Characterizing homozygous loss of function variants in 454,782 whole exome sequences from the UK Biobank
PrgmNr 3167 Conditional resampling improves calibration and sensitivity in single-cell CRISPR screen analysis
PrgmNr 3168 COVID19 symptoms vary by HLA haplogroup in a well typed crowd cohort
PrgmNr 3169 Deconvoluting sex-specific effects using GWAS summary statistics and biobank datasets across
PrgmNr 3170 Developing Trans-ethnic Polygenic Risk Scores Using Empirical Bayes and Super Learning Algorithm
PrgmNr 3171 Ethnicity-specific high-risk gene variant profiling unmasks diabetes associated genes
PrgmNr 3172 Exact and flexible, conditional and joint analysis of association summary statistics using linkage disequilibrium
PrgmNr 3173 Fast and scalable polygenic risk modeling with Variational Inference
PrgmNr 3174 Fine-mapping robustness in real data

PrgmNr 3175 Functional Response Regression Model on Correlated Longitudinal Microbiome Sequencing Data
PrgmNr 3176 Genealogical models of LD for Bayesian applications in GWAS
PrgmNr 3177 Genetic association analysis of a binary trait detects more than just the genetic effect: implications for pleiotropy and replication studies
PrgmNr 3178 GenoToolKit: A convenient and accessible R package for genetic data preparation and analysis
PrgmNr 3180 Novel Bayesian TWAS method using eQTL summary statistics
PrgmNr 3181 Polygenic risk prediction using gradient boosted trees captures non-linear genetic effects and allele interactions in complex phenotypes
PrgmNr 3182 Polygenic transcriptome risk scores improve cross-ethnic portability for COPD and lung function in the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program
PrgmNr 3183 Predicting overweight and obese with genome-wide, epigenome-wide gene-gene and gene-diet interactions using machine learning
PrgmNr 3184 Prioritizing disease genes by integrating GWAS and WES with gene-level functional and perturbation data
PrgmNr 3185 Rare variant effect sizes are associated with variant pathogenicity
PrgmNr 3186 Reproducible analyses of petabytes of Heart Lung Blood and Sleep (HLBS) data in the cloud using NHLBI BioData Catalyst, the data analysis ecosystem for HLBS researchers
PrgmNr 3187 Single-cell allele-specific expression analysis and comparison using mixture models
PrgmNr 3188 Sister haplotype association based on recombination disequilibrium
PrgmNr 3189 SUMMIT: An integrative approach for better transcriptomic data imputation improves causal gene identification
PrgmNr 3192 TIGAR-V2: Efficient TWAS Tool with Nonparametric Bayesian eQTL Weights of 49 Tissue Types from GTEx V8
PrgmNr 3194 Trans-Ancestry Fine-Mapping using 3.4 Million Individuals from diverse ancestries elucidates the genetic architecture of tobacco and alcohol use and addiction
PrgmNr 3195 Using (Natural) Direct and Indirect Effects of Intermediate Risk Factors between Genotype and Disease to Inform Polygenic Risk Scores
PrgmNr 3196 Allele-specific expression of SNPs involved in the immune response and gene transcription is associated with BMI
PrgmNr 3197 Causal gene-to-trait effect estimation for schizophrenia and bipolar disorder using MRLocus
PrgmNr 3198 Common and rare genetic risk for autism have opposing effects on cognition at autism-associated genes
PrgmNr 3199 Comparing Gene Expression Across Paired Human Airway Models for Cystic Fibrosis Precision Medicine
PrgmNr 3200 Current methods integrating variant functional annotation scores have limited capacity to improve the power of genome-wide association studies
PrgmNr 3201 Determining disease co-occurrence architecture of Hypertensive Heart Disease in Penn Medicine Biobank using longitudinal EHR data linked with PMBB participants
PrgmNr 3202 Genome-wide association study and heritability of Gulf War illness
PrgmNr 3203 Genome-wide study on longitudinal MRI outcomes in 3,513 multiple sclerosis cases across six clinical trials highlights a potential role for MS susceptibility loci in progressive biology
PrgmNr 3204 High quality phased whole genome sequence across patient cohorts is achievable and informs genetic understanding of complex traits
PrgmNr 3205 Leveraging Genetics and Genomics Data to Identify Potential Transcriptional Pathways of Body Mass Index: The Framingham Heart Study
PrgmNr 3206 Major sex differences in allele frequency for X-chromosome variants in the 1000 Genomes phase 3 data
PrgmNr 3207 Maternal and parent-of-origin gene-environment effects on the etiology of orofacial clefting

PrgmNr 3208 Mining biopsychosocial attributes in major depressive disorder by a multi-modal latent topic model

PrgmNr 3209 Patterns of consanguinity and associations between autozygosity and clinical phenotypes in British South Asians

PrgmNr 3210 Prioritizing risk factors of heart failure from UK Biobank using explainable artificial intelligence

PrgmNr 3211 Stability of polygenic scores across discovery genome-wide association studies

PrgmNr 3212 Systematic Comparisons for Composition Profiles Taxonomic Levels and Machine Learning Methods for Microbiome based Disease Prediction

PrgmNr 3213 Two large-scale genome sequencing studies show a consistent association between mitochondrial DNA copy number and personality traits

PrgmNr 3214 Unexpected somatic mosaicism and reduced penetrance: NF1 pathogenic variants are found in 1 in 850 individuals

PrgmNr 3215 Variants in nicotinic acetylcholine receptor and dopaminergic genes in relation to smoking relapse and proportion of time in relapse throughout adulthood in female nurses

PrgmNr 3216 NUDT15 Variants Associated with Thiopurine Toxicity in 1,643 Pediatric Leukemia Patients

PrgmNr 3217 A case report of rare t(8;22)(p11.2;q13) positive AML

PrgmNr 3218 A case study of AML in a patient with an additional RUNX1T1 signal and a loss of one 3' CBFB signal in 88% of the nuclei examined

PrgmNr 3219 A rare inv(14)(q11.2q32.3) in Acute Lymphoblastic Leukemia

PrgmNr 3220 Does it make sense to report single pathogenic alleles in cancer susceptibility genes for childhood cancer patients? Results from the BASIC3 and Texas KidsCanSeq studies

PrgmNr 3221 Heterogeneity in cytological presentation of gene amplification and concurrent amplification of MYC and PVT1 in therapy related acute myeloid leukemia (AML)

PrgmNr 3222 Imputing missing genotypes and examining rare genetic variants in families with high risk of lung cancer

PrgmNr 3223 Serial cytogenetic studies showing evolution of trisomy 8 positive CMML to AML with a complex karyotype characterized by deletion 7q and a jumping 1q translocation without evidence of trisomy 8

PrgmNr 3224 Somatic Mutation Profiling of Basal Cell Carcinoma in a Population Exposed to Arsenic

PrgmNr 3225 The g-quadruplex stabilizing agent GQC-05 inhibits LINE-1 retroelement activity in human bronchial epithelial cells

PrgmNr 3226 Underutilization of germline testing for prostate cancer patients: Are genetic testing criteria a tool or an obstacle?

PrgmNr 3227 Age-specific transcriptional risk scores (TRS) link GWAS to eQTLs and predict therapeutic response across 8 common cancer types

PrgmNr 3228 Broadly pleiotropic effects of pathogenic variants in tumor suppressor CHEK2

PrgmNr 3229 Genome-wide interaction analysis identified low-frequency variants with sex disparity in lung cancer risk

PrgmNr 3230 GWAS of Uveal Melanoma Reveals Novel Genome Wide Significant Locus

PrgmNr 3231 Predicted gene expression identifies novel genes associated with lung cancer in African Americans

PrgmNr 3232 SNPs at SMG7 associated with time from biochemical recurrence to prostate cancer death

PrgmNr 3233 Cisplatin induced gene expression across three different cancer cell line types reveal common and distinct molecular regulatory effects

PrgmNr 3234 Computational evidence of BRCA1-NC3 receptor interaction and druggability analysis in BRCA1-related oncogenesis

PrgmNr 3235 Identifying functional melanoma risk variants and susceptibility genes via massively parallel reporter assays and multi-QTL in normal melanocytes and malignant melanomas

PrgmNr 3236 A deep learning approach for efficiently incorporating genomic data into ovarian cancer prognosis prediction

PrgmNr 3237 An improved approach leveraging MRI and genomic information for classification and segmentation of glioma lesions
PrgmNr 3238 Analyses of breast cancer microbiome data
PrgmNr 3239 Building multi-omics classifier to improve phenotype prediction in heterogeneous cancer data
PrgmNr 3240 Characterization and prediction of DNA methylation instability across human cancers
PrgmNr 3241 Diagnosing Li-Fraumeni syndrome from the somatic genome
PrgmNr 3242 Epigenomic landscape and 3D genome structure in pediatric high-grade glioma
PrgmNr 3243 MDC1 restrains ATR-mediated resection of DNA double-strand breaks in human cells
PrgmNr 3244 Single cell transcriptome analysis maps aneuploidy to mesenchymal phenotypes during thyroid cancer progression
PrgmNr 3245 Statistical analysis of breast cancer co-methylation patterns
PrgmNr 3246 Subtype-associated epigenomic landscape and 3D genome structure in bladder cancer
PrgmNr 3247 Survival analysis and evaluation of germline genomic patterns across multiple cancers using boosted trees and model interpretability
PrgmNr 3248 The clinical impact of somatic variants in AGT, MGMT and TP53 in Mexican patients with astrocytoma
PrgmNr 3249 Tumor spectrum of FLCN gene mutation carriers in an unselected pan-cancer patient population
PrgmNr 3250 Up-regulation of chromatin remodeling genes contributes to chemoresistance and tumorigenesis in PDAC
PrgmNr 3251 Assessment of population-appropriate polygenic risk scores for lipid traits in African Americans
PrgmNr 3252 Bivariate genome-wide association study of circulating fibrinogen and C-reactive protein (CRP) levels
PrgmNr 3253 Blood Pressure Polygenic Score Associations in a Pediatric Cohort Requiring Surgery for Congenital Heart Defects
PrgmNr 3254 Detecting blood-pressure regulatory mechanisms with tissue specific CRE maps and biobank scale datasets
PrgmNr 3255 Epigenetic age acceleration predicts subclinical atherosclerosis among participants of the Bogalusa Heart Study
PrgmNr 3257 Identification of genetic variants that confer susceptibility to familial combined hyperlipidemia
PrgmNr 3258 Phenotypic features of TGFB3 disease-causing variants in a group of North Americans
PrgmNr 3259 Polygenic risk for atrial fibrillation across ancestry: the Electronic Medical Records and Genomics (eMERGE) Network
PrgmNr 3260 Polygenic risk score for coronary artery disease worsens disparity by HIV status in clinical prediction of coronary artery calcium
PrgmNr 3261 The genetic determinants of body mass index & fasting glucose are mediators of grade 1 diastolic dysfunction development
PrgmNr 3262 The genetics of coronary artery calcification in individuals with type 2 diabetes
PrgmNr 3263 An exposome-wide association study (ExWAS) and comparison of clinical, environmental and polygenic risk scores for Type 2 Diabetes
PrgmNr 3264 Assessment of Polygenic Risk Score for Predicting Type 2 Diabetes Risk in Asian Indians: Results of the Asian Indian Diabetic Heart Study/Sikh Diabetes Study
PrgmNr 3265 Do methods of inclusion and presentation of non-European ancestry data in GWAS allow for novel discoveries?
PrgmNr 3266 Expanded Genetic Clustering of Type 2 Diabetes Loci Using a High-throughput Pipeline Reveals Mechanistic Pathways of Metabolic Diseases
PrgmNr 3267 Genetic Regulation of Obesity Explains Cardiovascular Sex Differences in Polycystic Ovary Syndrome for Predisposed Individuals
PrgmNr 3268 Genotype-Phenotype Analysis in African Americans with Inflammatory Bowel Disease

PrgmNr 3269 Identification of hepatocellular gene expression phenotypes influencing fatty liver disease in Hispanics

PrgmNr 3270 Investigating the Causal Role of Reduced Vitamin D Levels With Type 2 Diabetes Risk in South Asians and Europeans

PrgmNr 3271 Pediatrics and proteins: Molecular discovery in the Hearts & Parks study of children with obesity

PrgmNr 3272 Phenome-wide investigation of health outcomes associated with the genetic correlates of 25 hydroxyvitamin D concentration

PrgmNr 3274 Rare variants associated with serum creatinine in 350k UK Biobank whole exome sequences

PrgmNr 3275 The effect of obesity-related traits on COVID-19 severe respiratory symptoms and hospitalization and its mediation by socioeconomic status: a multivariable Mendelian randomization study

PrgmNr 3276 Whole exome sequencing analyses of BMI and obesity in a multi-ancestry cohort of > 300,000 individuals

PrgmNr 3277 Whole Genome Sequence Analysis of Dyspnea in Current and Former Smokers in COPDGene

PrgmNr 3278 Association of CXCR6 with COVID-19 severity: host genetic factors underlying transcriptomic regulation in 3p21.31 locus

PrgmNr 3279 Asthma-associated genetic variants induce IL33 differential expression through a novel regulatory region

PrgmNr 3280 Complex Architecture of COVID-19 Host Genetics from Whole Genome Sequencing of a Multi-Ethnic Hospitalized Cohort

PrgmNr 3281 Fine-mapping studies distinguish genetic risks for childhood- and adult-onset asthma in the HLA Region

PrgmNr 3282 Infinity Biologix - A Biorepository Answers the Call and Becomes a Leading COVID-19 Assay Development and Testing Center

PrgmNr 3283 Integrative analysis of DNase-I footprinting and genome-wide association data uncovers transcription factors involved in nine autoimmune diseases

PrgmNr 3284 MHC class II alleles are associated with rare dermatologic autoimmune diseases in an analysis of over 200,000 participants of the UK Biobank

PrgmNr 3285 mRNA expression analysis reveals degenerative phenotype and immune cell infiltration in spleen, thymus and pancreas of Clec16a knockout (KO) mice

PrgmNr 3286 Role of theFYVE and Coiled-Coil Domain Autophagy Adaptor 1 in severity of COVID-19 infection

PrgmNr 3287 The long-term effect of SARS-CoV-2 infection on host genetic regulation

PrgmNr 3288 Trans-ethnic association analysis of HIV viral load and Chromosome X markers in individuals living with HIV

PrgmNr 3289 Whole-genome sequencing association study reveals methylation-mediated genotype association with decreased lung function in urban children

PrgmNr 3290 APOL1 and biobanking in the West African terrain-challenges and successes

PrgmNr 3291 A Myasthenia Gravis genomewide association study implicates AGRN as a risk locus

PrgmNr 3292 Cognitive trajectories diverge by genetic risk in the Wisconsin Registry for Alzheimer's Prevention

PrgmNr 3293 Concordance of genetic variation that affects neuroinflammatory biomarkers for Alzheimer's disease and that influences brain volumes

PrgmNr 3294 Creating sex-specific phenome risk classifiers to identify under-documented cases of developmental stuttering in electronic health records

PrgmNr 3295 Genetic correlations between resilience and UK Biobank traits differ by biological sex

PrgmNr 3296 Genetic variants in the SHISA6 gene are associated with delayed cognitive impairment in two family datasets

PrgmNr 3297 Genomic and Transcriptomic-wide Analysis Identifies Novel Genetic Risk Loci and Prioritization of Therapies for Myasthenia Gravis

PrgmNr 3298 Identification of novel genetic associations of epilepsy with functional validation in a zebrafish model

PrgmNr 3299 Linkage and association of cognitive resilience to chromosomes 4 and 18 in Midwestern Amish

PrgmNr 3300 Measuring similarities and differences between EOAD and LOAD trough genome-wide association study, colocalization and heritability analysis

PrgmNr 3301 APOE gene associated with dementia related disorders and neuropsychiatric diseases in the Hispanic population

PrgmNr 3302 Association between polygenic risk for substance use and music engagement across multiple cohorts

PrgmNr 3303 Dimensional traits in genetic research of autism show greater heritability and generalization than case-control traits

PrgmNr 3304 Identify high risk individuals for Alcohol use disorder using polygenic risk scores

PrgmNr 3305 Identifying potential environmental influences on clinical comorbidities of schizophrenia through integration of electronic health data and genetics

PrgmNr 3306 Investigating shared genetic liability between recurrent major depression and cardiometabolic traits in East Asian populations

PrgmNr 3307 Pilot targeted gene analysis of resting heart rate variability in individuals with anxiety disorders and healthy controls

PrgmNr 3308 Targeted sequencing of genes most likely to harbor rare causal variants in schizophrenia

PrgmNr 3309 MS4A gene cluster variant modifies CSF sTREM2 associations with AD biomarkers of pathology

PrgmNr 3310 Analysis of genome-wide associations, exposome and lifestyle factors of skin-aging using medical proxies in the UK BioBank

PrgmNr 3311 Associated phenotypic and genotypic characteristics in a cohort of patients with infantile hemangiomas: a case series

PrgmNr 3312 Association of whole-person polygenic component scores with Alzheimer's disease pathology

PrgmNr 3313 Candidate causal variants affecting tissue-specific regulatory mechanisms in the GWAS catalog

PrgmNr 3314 Curated catalog of 309 functional GWAS non-coding variants across 125 diseases: a systematic review

PrgmNr 3315 Determining the Genetic risk of Complex Disorders in Global populations

PrgmNr 3316 DNA-methylation quantitative trait loci in Asian populations

PrgmNr 3317 Fibromyalgia as a Nexus of Multiple Liability Distributions

PrgmNr 3318 Genetic architecture of risky sexual behavior: Correlational structure and causal effects

PrgmNr 3319 Genome-wide association study of anthropometric traits in a health-disparate Hispanic population

PrgmNr 3320 Genomic approaches to identify shared genetic architecture among comorbid phenomes of eye disease

PrgmNr 3321 Individual Imputation of Genetically Regulated Gene Expression in the Million Veteran Program

PrgmNr 3322 Investigation of a complex phenotype involving craniosynostosis and cataracts leads to identification of a novel mutation related to Galactosemia type II

PrgmNr 3323 Modeling the longitudinal changes of ancestry diversity in the MVP cohort: systematic comparison of methods

PrgmNr 3324 Novel Osteoarthritis Genetic Variants Identified using Multi-Ancestry Analyses from 473,388 Participants in the MVP and the UK Biobank

PrgmNr 3325 Novel phenotyping approaches: Empirical Bayes methods improve heritability of dietary phenotypes from repeat diet questionnaire data

PrgmNr 3327 Programming biobank-scale relatedness analysis improvements

PrgmNr 3328 Reanalysis of exome data via Diagnosticator identifies significant variant reclassification with updated ClinGen SVI-WG recommendations
PrgmNr 3329 Study of over one million people finds that novel and previously reported familial genetic effects impact stuttering risk at a population-level
PrgmNr 3330 The contribution of common versus low-frequency variation to hearing loss heritability
PrgmNr 3331 Transcriptome-wide association study and eQTL colocalization identify potentially causal genes responsible for bone mineral density GWAS associations
PrgmNr 3332 Whole-exome sequencing of a family with specific language impairment suggests BUD13 as a novel SLI candidate gene
PrgmNr 3333 Year of birth and polygenic scores in the Electronic Health Records
PrgmNr 3334 A global transcriptomic approach reveals sex-specific differences in a middle-aged frailty cohort
PrgmNr 3335 A single cell chromatin accessible and transcriptome atlas of the human heart facilitates identification of risk variants and genes of Atrial Fibrillation
PrgmNr 3336 Blood-based differential DNA methylation and gene expression at the TREM locus in mild traumatic brain injury and Alzheimer's disease
PrgmNr 3337 Comparing gene expression imputation resources in a diverse Latin American Parkinson's disease cohort
PrgmNr 3338 Cross-omics analysis of DNA methylation and gene expression in post-mortem 41 brain tissues of opioid use disorder patients and controls
PrgmNr 3339 Identifying and characterizing superhypomethylation in head and neck cancer: drivers and transcriptional effects
PrgmNr 3340 Integrative meta-analysis of blood and brain epigenome-wide association studies of Alzheimer's disease
PrgmNr 3341 Methylation Profiling using Reduced Representation Bisulfite Sequencing for Basal Cell Carcinoma in a Population Exposed to Arsenic
PrgmNr 3342 Methylome-Wide Association Study Identifies Candidate DMRs for Carotid Bifurcation Intima-Media Thickness in Dominican Republic Families
PrgmNr 3344 PTSD, major depression, and advanced transcriptomic age in neural tissue
PrgmNr 3345 Role of differential DNA methylation in Mexican Americans & their risk for Alzheimer's disease & type 2 diabetes
PrgmNr 3348 Transcriptomic and epigenomic signatures associated with autoimmunity in Parkinson's Disease
PrgmNr 3349 A cryptic splice acceptor as a cause of X-linked heterotaxy
PrgmNr 3350 Adipose eQTL meta-analysis of 2,256 samples identifies 517 conditionally distinct signals colocalized with 426 cardiometabolic trait GWAS loci
PrgmNr 3351 Characterization of neonatal epigenetic markers associated with elevated stress from the COVID-19 pandemic
PrgmNr 3352 Defining the effects of noncoding genetic variation on human regulatory element activity
PrgmNr 3353 DNA-binding consequences of human homeodomain missense variants
PrgmNr 3354 Genome-wide association studies identify novel genetic loci for epigenetic age acceleration among survivors of childhood cancer
PrgmNr 3355 Genome-wide detection of enhancer-hijacking and reconstructing complex SVs in cancer genome
PrgmNr 3356 Automated localization and quantification of RNA transcripts from RNA-fluorescent in situ hybridization imaging data
PrgmNr 3357 Deep Learning Study to Predict Ribosome Binding from RNA-seq Data
PrgmNr 3358 Interpreting Epigenetic Aging with Deep Neural Networks Using High Dimensional DNA Methylation Data
PrgmNr 3359 Investigating the utility of sex-informed PrediXcan models in whole blood
PrgmNr 3360 A site-specific mutation rate model uncovers signatures of localized hypermutability

PrgmNr 3361 Assessment of the rates of pathogenic variants across the All of Us Research Program cohort

PrgmNr 3362 Cataloging human PRDM9 allelic variation using long-read sequencing reveals PRDM9 population-specificity and two distinct groupings of related alleles

PrgmNr 3363 Characterization of the Evolutionary Forces Acting on Genomic Regions Associated with over 600 Human Complex Traits

PrgmNr 3364 Core promoter similarity in ancient and modern human and in chimpanzee genomes

PrgmNr 3365 Estimating average height among populations from polygenic risk scores

PrgmNr 3366 FastRecomb: Fast inference of genetic recombination rates in biobank scale data using positional Burrows Wheeler transform

PrgmNr 3367 Flexible, scalable inference with tree sequences and Pyro

PrgmNr 3368 Guaranteeing unbiasedness in selection tests based on polygenic scores

PrgmNr 3369 GWAS on birth year infant mortality rates provides new evidence of recent natural selection

PrgmNr 3370 Increasing Portability of Predictive Gene Expression Models Across Ancestral Populations Using Genetic Variant Fixation Indices

PrgmNr 3372 The genetic diversity reduction due to background selection increases the disease prevalence under the liability threshold model

PrgmNr 3373 A comprehensive genomic test report for communicating cancer and incidental findings

PrgmNr 3374 A novel model for the management of clinically relevant incidental findings

PrgmNr 3375 Addressing the Direct-to-Consumer Genetic Testing Knowledge Gap for Non-Genetics Healthcare Professionals

PrgmNr 3376 BorderDNA Resources Project: Strategies for the development and disseminationof educational materials on nonmedical DNA applications in immigration contexts

PrgmNr 3377 Evaluation of genetic services in the care for adults with tetralogy of Fallot

PrgmNr 3378 Exploring the Use of Family Health History in Black Communities in the Deep South

PrgmNr 3379 Genetic service provision for men with prostate cancer within a safety-net setting: A qualitative study

PrgmNr 3380 Genetic-first diagnoses and disease predispositions: Implications for follow-up care

PrgmNr 3381 Grandmothers save the day: The power of grandmothers to improve care of families with history of cardiomyopathy

PrgmNr 3382 Integrating Ethical, Legal, and Social Implications (ELSI) considerations in the All of Us Research Program:Experiences of the ELSI Brain Trust

PrgmNr 3383 Measuring genetic literacy through familiarity, skills, and knowledge: A longitudinal assessment

PrgmNr 3384 Patient perspective on return of incidental findings in Africa

PrgmNr 3385 Patients demonstrate high motivation to enhanced cancer screening following genetic counselling

PrgmNr 3386 Recruitment of diverse participants into a trial of polygenic risk scoring: Enrollment metrics and reasons for decline in the Genomic Medicine at VA (GenoVA) Study

PrgmNr 3388 Should laboratories report small regions of homozygosity?

PrgmNr 3389 Strategies to increase clinicians in medical genetics: Establishing a pipeline to attract and prepare diverse clinical researchers

PrgmNr 3390 The Texome Project: Special considerations for developing a rare and undiagnosed disease program for underserved communities in Texas

PrgmNr 3392 Uptake of cancer risk management strategies among women who undergo cascade genetic testing for breast cancer susceptibility: How do they compare to non-cascade testers?

PrgmNr 3393 Colchicine: A Potential Therapy for the HSPB8-related Rimmed Vacuole Myopathy

PrgmNr 3394 Genome-wide integration assay for rAAV mediated homologous recombination (HR) in human hepatocytes demonstrated precision of in vivo gene editing approach

PrgmNr 3395 A New Paradigm for Coordinate-Based, Multipanel Genomic Plotting in R

PrgmNr 3396 All of Us Researcher Workbench: Bringing Researchers to Genomics Data

PrgmNr 3397 Annotation of genetic variation in 5134 cystic fibrosis cases with OpenCRAVAT and VEP

PrgmNr 3398 FusionGDB 2.0, fusion gene annotation update aided by deep learning

PrgmNr 3400 Genomics4RD: An integrated platform to share Canadian deep-phenotype and variant level data for international rare disease gene discovery

PrgmNr 3401 HGRepo - A comprehensive repository of clinical, lab assay, and genetic data for studies of innate and adaptive immune responses during acute COVID-19 infection and convalescence

PrgmNr 3402 New genomic data deposited in dbGaP cross-referenced against NINDS Repository Tourette Syndrome biospecimens

PrgmNr 3403 A Finnish-enriched functional DNA variant in an enhancer for IRF6 is associated with cleft palate

PrgmNr 3404 Analysis of genetic investigations in newborns with VACTERL spectrum of malformations

PrgmNr 3405 Biallelic variants in KARS1 are associated with neurodevelopmental disorders and hearing loss recapitulated by the knockout zebrafish

PrgmNr 3406 Congenital heart defects caused by FOXJ1

PrgmNr 3407 Loss of Zic3 impairs planar cell polarity leading to abnormal left-right signaling and heart defects

PrgmNr 3408 Novel pathogenic variants, evidence for involvement of DVL2, and quantitative phenotypic analyses of Robinow syndrome: WNT signaling perturbation and phenotypic variability

PrgmNr 3409 De novo TNPO2 variants are associated with developmental delays, neurological issues and dysmorphic features in humans and alter TNPO2 activity in Drosophila

PrgmNr 3410 De novo damaging variants in the microRNA processor DROSHA are associated with a severe progressive neurological disorder

PrgmNr 3411 De novo variant in MRTF-B is associated with intellectual disability, minor dysmorphic features, expressive language delay, impulse control issues, and fine motor delay

PrgmNr 3412 SLC6A1 Portal: Genetic variant analysis and educational resource for SLC6A1-related disorders

PrgmNr 3414 A novel de novo variant in PPP2CA affecting the 5'UTR region causes a syndrome characterized by global developmental delay with specific structural brain abnormalities, congenital microcephaly and dysmorphic features

PrgmNr 3415 Allelic frequency and distribution of M98Kvariant associated with adult-onset open angle glaucoma in Puerto Rico

PrgmNr 3416 An integrated machine learning and functional analysis approach for resolution of variants of uncertain significance (VUS) inSTXBP1

PrgmNr 3417 Biallelic NAV2 truncating variants cause a neurodevelopmental disorder with cerebellar cortical dysplasia

PrgmNr 3418 Delineation of a neurodevelopmental syndrome caused by PAX5 haploinsufficiency

PrgmNr 3419 Diagnostic rate and clinical implications of exome sequencing in adults with intellectual disability

PrgmNr 3420 Expanding the Genotypic and Phenotypic Spectrum of CTCF-Related Disorders: 19 new patients andreviews of published literature

PrgmNr 3421 Expanding the phenotypic spectrum of EXOSC8-related neurological disease: Presentation of an Egyptian family with neurodevelopmental disorder due to a novel EXOSC8 missense variant

PrgmNr 3422 Genetic heterogeneity in SARS2 gene in a patient with global developmental delay and spastic paraplegia

PrgmNr 3423 Genomics of early onset CNS autoimmune demyelinating diseases: Role for deleterious, rare variants in common immune pathways

PrgmNr 3424 Identifying variant specific phenotypes for SLC6A1 related disorders

PrgmNr 3425 KBG Syndrome: Prospective Videoconferencing and Use of AI Facial Recognition in 25 new patients

PrgmNr 3426 Loss of function variants in PUS7 dysregulate protein synthesis and are associated with neurodevelopmental delay, microcephaly, and hyperuricemia

PrgmNr 3427 Maternal mosaicism for a missense variant in the SMS gene that causes Snyder-Robinson syndrome

PrgmNr 3428 Novel PRUNE1 A (p.Thr311=) synonymous splice variant induces exon 7 skipping and leads to an atypical presentation of NMIHBA syndrome: Case report and review of the literature

PrgmNr 3429 Role of VRK1 defects in spinal muscular atrophies like patients with negative SMN1 finding

PrgmNr 3430 Role of exome testing in autism spectrum disorder

PrgmNr 3431 The phenotypic trajectory of 16p12.1 deletion is determined by family history of disease

PrgmNr 3432 The Use of Acetazolamide for Apneic Episodes in a Patient with EEF1A2 Mutation: A Case Report

PrgmNr 3433 Triple knockout ofFmr1,Fxr1andFxr2in Purkinje neurons results in mild behavioral phenotypes and a dramatic reduction of testis weight

PrgmNr 3434 Variation of Serum Metabolome and Exosomal Transcriptome in Acute Ischemic Stroke: Results of the Metabolome in Ischemic Stroke Study (MISS)

PrgmNr 3435 A heterozygous gain-of-function variant in KIF5B causes osteogenesis imperfecta by disrupting the Golgi-primary cilia axis

PrgmNr 3436 Biallelic variants in LDB3: A new inheritance pattern for an old cardiomyopathy gene

PrgmNr 3437 Combined PMM2-CDG and hereditary fructose intolerance in a child with mild clinical presentation: a case report

PrgmNr 3438 Implementing genome and transcriptome sequencing methods to improve the diagnosis of Mendelian myopathies

PrgmNr 3439 Intravenous fish oil emulsion and fatty acid profiles

PrgmNr 3440 Lessons learned from a severe Isovaleric acidemia (IVA) patient from hyperammonemia

PrgmNr 3441 Mutational spectrum in a sample of mexican patients with Hunter syndrome

PrgmNr 3442 Mutations in COL5A1 lead to heritable thoracic aortic disease

PrgmNr 3443 Oculocutaneous Albinism: Compound heterozygote genotype-phenotype correlation

PrgmNr 3444 Sengers syndrome: Novel pathogenic variant in the AGK gene and analysis of phenotypic variability and clinical outcomes

PrgmNr 3445 Whole genome sequencing of mitochondria: A powerful tool for clinical diagnosis of human mitochondria diseases

PrgmNr 3446 EFEMP1 rare variants cause juvenile-onset open angle glaucoma in families from the Philippines

PrgmNr 3447 FOXP1 haploinsufficiency contributes to the development of congenital diaphragmatic hernia

PrgmNr 3448 A multidisciplinary nephrogenetic referral clinic for children and adults - diagnostic achievements and insights

PrgmNr 3449 Carrier Prevalence of 302 genetic diseases in the Mexican Jewish Community

PrgmNr 3450 Congenital diarrhea is a common cause of neonatal crisis and up to ~60% of cases among neonates may be inherited

PrgmNr 3451 Diagnostic rate of clinical exome sequencing for CAKUT phenotypes

PrgmNr 3452 Exome sequencing further implicates disruption of the NODAL signaling pathway in laterality defects

PrgmNr 3453 Identifying susceptibility genes for congenital diaphragmatic hernia (CDH) using a machine learning algorithm

PrgmNr 3454 Increased Expression of ZFPM2 Bypasses SRY to Drive 46,XX Testicular Development: A New Mechanism of 46,XX DSD

PrgmNr 3455 Patient with novel compound heterozygous variants in NCKAP1L presents Immunodeficiency 72 with autoinflammation and novel features

PrgmNr 3456 Precision medicine models for undiagnosed and rare disease

PrgmNr 3457 Rapid CRISPR/Cas9-based method for functional validation of human disease genes

PrgmNr 3458 Refined SNP haplotype analysis uncovers a potential founder deletion mutation in the Cerebral Cavernous Malformations 2 gene

PrgmNr 3459 Significant burden of de novo damaging variants in novel genes in patients with congenital kidney malformations

PrgmNr 3460 Complex congenital heart defect in a patient with a pathogenic mutation of TLK2 gene: Coincidental findings or part of the phenotype of TLK2 syndrome?

PrgmNr 3461 Partial trisomy of 8q and partial monosomy of 10q as a result of unbalanced translocation in an infant with multiple congenital anomalies - A Case Report

PrgmNr 3462 A rapid and low cost NGS-based cytogenomics platform using proximity ligation technology

PrgmNr 3463 A Retrospective Review of Dominant Variants Identified from the HorizonTM Carrier Screening Test

PrgmNr 3464 Contiguous domains of differentially-accessible, single copy sequences along metaphase chromosomes are conserved among multiple tissues

PrgmNr 3465 Detection of near constitutional genome-wide uniparental disomy in a patient presenting with Beckwith-Wiedemann syndrome and cholestasis

PrgmNr 3466 Inverted genomic triplication structures: two breakpoint junctions, several possibilities

PrgmNr 3467 Library preparation for DNAseq: The next generation of automation

PrgmNr 3468 A novel intragenic deletion removing the secondary coiled-coil (CC2) domain of SOX6 in a patient with neurodevelopmental disorder

PrgmNr 3469 Analysis of single nucleus sequencing in rat amygdala reveals transcriptional and regulatory mechanisms associated with vulnerability to cocaine addiction

PrgmNr 3470 Arimoclomol's effects on Rimmed Vacuole myopathy caused by HSPB8 gene mutation

PrgmNr 3471 Characterization of APOE Christchurch carriers in 239,390 participants of the UK Biobank with whole exome and/or whole genome sequencing data

PrgmNr 3472 Drosophila functional screening of de novo variants in autism uncovers deleterious variants and facilitates discovery of rare neurodevelopmental diseases

PrgmNr 3473 Evaluation of chromosomal copy numbers in the 5'UTR of MBD5

PrgmNr 3474 Functional characterization of 400 acid α-glucosidase variants and analysis of phenotype-genotype relationships in Pompe disease

PrgmNr 3475 Genetic regulation of gene co-expression in human prefrontal cortex

PrgmNr 3476 Identifying novel evolutionarily conserved regulators of dopamine metabolism using a forward genetic screen in Drosophila melanogaster

PrgmNr 3477 Interrogating strategies to link CRISPR screens to neurodegenerative disease

PrgmNr 3478 Lessons learned from genomics studies of a large recessive NDD cohort: High prevalence of multilocus pathogenic variation, candidate novel genes, and extreme genetic heterogeneity

PrgmNr 3480 Molecular Dynamics Simulations and Markov State Models Reveal Metastable States and Isoform-Dependent Kinetics in Apolipoprotein E Risk and Resilience Alleles

PrgmNr 3481 Mutations Near the N- or C-Terminus of AHDC1 are Associated with More Mild Xia-Gibbs Syndrome Phenotypes

PrgmNr 3482 Personalized structural biology reveals the molecular mechanisms underlying heterogeneous epileptic phenotypes caused by de novo KCNC2 variants

PrgmNr 3483 MPDZ, a potential causal gene for the bone mineral density association at chromosome locus 9p23, regulates in vitro osteoblast function and in vivo bone mass accrual

PrgmNr 3484 A Novel Role for HAPLN1 in Skeletal Development

PrgmNr 3485 A SNP associated with risk for non-syndromic orofacial cleft affects expression of FOXE1 in an oral epithelium cell line

PrgmNr 3486 Clinical application of long-read sequencing technology for unsolved cases of neurodevelopmental disorders

PrgmNr 3487 Functional interpretation of TAB2 missense variants

PrgmNr 3488 Germline variation in the human immunoglobulin heavy chain locus contributes to inter-individual variability in the expressed antibody repertoire

PrgmNr 3489 High Throughput Trafficking Assay Identifies Novel Loss and Gain of Function Variants in the Potassium Channel Gene KCNE1

PrgmNr 3491 Hyperproliferative lymphatic sprouting caused by an NRAS mutation can be rescued by targeting PI3K/AKT and RAS/MAPK signaling pathways in cell-based modeling systems

PrgmNr 3492 Seeking answers for ocular disease: Characterizing intronic variants of uncertain significance using a minigene system

PrgmNr 3493 Single-cell RNA sequencing of red and black melanocytes uncovers new pathways in pigmentation and disease

PrgmNr 3494 A computational approach for detecting physiological homogeneity in the midst of genetic heterogeneity

PrgmNr 3495 A statistical analysis framework to test the presence of allelic imbalance from a genomic region using allelic count data

PrgmNr 3496 Assessing Short-Read Utility for SVs

PrgmNr 3497 Copy-Number Variation of CTCF Binding Sites Associates with Diverse Clinical Phenotypes

PrgmNr 3498 Evaluation of long-reads across challenging medical relevant genes and its implications for All of Us

PrgmNr 3499 Framework and analysis of cell-specific expression contributions to Mendelian Disease

PrgmNr 3500 Haplotype resolved LPA and Kringle repeats to study minorities across the USA

PrgmNr 3501 Identification of allele-specific KIV-2 repeats among multi-ethnic groups and association with LP(a) measurements

PrgmNr 3502 Knowledge based artificial intelligence for variant pathogenicity prediction for Mendelian disorders

PrgmNr 3503 MetaRNN: Differentiating rare pathogenic and rare benign missense SNVs and InDels using deep learning

PrgmNr 3504 Participation of rs9138 and rs1126616 of SPP1 gene in gender-specific predisposition in the development of systemic lupus erythematosus

PrgmNr 3505 Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data

PrgmNr 3506 Resolving the landscape of human genetic variation at the protein domain level using 454,803 exomes from the UK Biobank

PrgmNr 3507 The Genetic Architecture of Genome-scale Metabolic Networks

PrgmNr 3508 The structural landscape of constrained sites in the human proteome

PrgmNr 3509 Transposable element mediated rearrangements are prevalent in human genomes

PrgmNr 3510 Ultra-fast genome-wide association analysis for Alzheimer's Disease in 200,000 exomes from UK Biobank

PrgmNr 3511 Unique Sequence Detection: a novel variant genotyper spanning the full spectrum of genetic variants

PrgmNr 3512 Analysis of concordance of genetic variants across sample types

PrgmNr 3513 Automating Illumina RNA Prep With Enrichment Using Beckman Coulter i7 Hybrid Automated Workstation

PrgmNr 3514 Automating Twist Bioscience Modular Library Preparation and Targeted Enrichment Protocols Using Beckman Coulter Automation

PrgmNr 3515 Automation of NGS library preparation: Generate high quality libraries with less hands-on time

PrgmNr 3516 Germline and somatic protein quantitative trait loci from the tumor proteome in high grade serous ovarian cancer identify candidate biomarkers and connections between germline risk and disease biology

PrgmNr 3517 Impact of DNA contaminants on next-generation library preparation

PrgmNr 3518 Integrated multi-omics analysis of thyroid cancer reveals key molecular pathways involved in tumor formation and metastasis

PrgmNr 3519 Mitigating effects of blood disease pathologies that compromise specificity of gene expression signatures for radiation exposure

PrgmNr 3520 Untargeted metabolomics on the diet behaviors and their genomic interaction among Mexican Americans

PrgmNr 3521 Best practices for trans-ethnic, meta-analytic transcriptome-wide associations: lessons from the Global Biobank Meta-Initiative

PrgmNr 3522 Accurate profiling of forensic autosomal STRs using the Oxford Nanopore Technologies MinION device

PrgmNr 3523 Development of a highly multiplexed RNA in situ hybridization assay for spatial transcriptomic mapping of formalin-fixed paraffin-embedded tissue

PrgmNr 3524 DRIVE: Distant relatedness for identification and variant evaluation

PrgmNr 3525 Enabling faster and more cost-effective drug discovery through flexFSTM: A highly durable, elastic, and scalable cloud-native file system

PrgmNr 3526 Enabling Real Time Analysis of Single-Cell Genomic Data with Accelerated Computing

PrgmNr 3527 Improving rare variant calls from genotype array data: the Million Veteran Program Return Of Actionable Results (MVP-ROAR) Study

PrgmNr 3528 Integration of WGS & RNA-seq as an efficient approach for novel variant discovery in complex genetic cases

PrgmNr 3529 Introduction of a walk-away automated Roche NGS workflow solution: Integrated KAPA library preparation, KAPA target enrichment and the AVENIO Edge instrument

PrgmNr 3530 Investigation of a dysmorphic facial phenotype in patients with Gaucher type 2 and 3

PrgmNr 3531 Multiple de novo copy number variant (MdnCNV) driven mirror traits and blended phenotype

PrgmNr 3532 Rapid endpoint PCR: Improving amplification speed without sacrificing specificity

PrgmNr 3533 Sequence-based assessment of mitochondrial DNA oxidative damage in cognitive impairment: Shedding light on health disparities in Mexican Americans

PrgmNr 3534 TowardsIsoform Resolution Single-Cell Transcriptomicsfor Clinical Applications Using Highly Accurate Long-Read Sequencing

PrgmNr 3535 Attitudes on pharmacogenetic results as secondary findings among medical genetics providers

PrgmNr 3536 Connectivity Map Analysis of a Single-Cell RNA-Sequencing Derived Transcriptional Signature of mTOR Signaling

PrgmNr 3537 Explainable machine learning for adverse drug reactions

PrgmNr 3538 Identifying the genetic contributors of efficacy and adverse metabolic effects of thiazide diuretics in African Americans from the Genetics of Hypertension Associated Treatments (GenHAT) study

PrgmNr 3539 Investigation of matrisome genes within the druggable genome

PrgmNr 3540 Long-read amplicon sequencing of the polymorphic CYP2D6 locus

PrgmNr 3541 Targeting Alzheimer’s Disease drug development via network analysis of GWAS data

PrgmNr 3542 Worse COVID-19 outcomes in patients with the hemoglobin beta sickle allele due to existing comorbidities and new acute events

PrgmNr 3543 A Single-Assay Solution for Expanded Carrier Screening Relieves Existing Workflow Constraints and Provides More Comprehensive Analysis

PrgmNr 3544 Cord blood DNA methylation gestational age estimation applied to dried blood spot DNA methylation

PrgmNr 3545 Developmental genomics of congenital limb malformations: further evidence for gene dosage effects

PrgmNr 3546 Potential additive effect of fetal HRAPOL1genotype and micronutrient deficiencies

PrgmNr 3547 A novel polygenic score for microglial activation enhances models of Alzheimer's disease neuropathology

PrgmNr 3548 A novel test for parent-of-origin effects in population samples leveraging multiple phenotypes

PrgmNr 3549 A powerful and resource-efficient pipeline for association analysis of large-scale whole-genome sequencing studies
PrgmNr 3550 A quantile integral linear model to quantify genetic effects on phenotypic variability
PrgmNr 3551 A resource-efficient tool for phenome-wide rare variant association analysis in large-scale whole-genome sequencing studies, with application to TOPMed metabolomics data
PrgmNr 3552 An extended fine-mapping framework to identify causal genes from transcriptome-wide association studies (TWAS)
PrgmNr 3553 An iterative approach to adjustment for covariates in genetic association studies using GWAS summary statistics
PrgmNr 3554 Ancestry-matching cases to controls from the GLAD database to accelerate genetic research in under-represented populations
PrgmNr 3555 Bayesian estimation of Allele-Specific Expression (ASE) in the presence of phasing uncertainty for identifying genes with aberrant expression
PrgmNr 3557 Consistent summary counts based rare variant analysis for discovery of germline predisposition genes using public genotype summary data as controls
PrgmNr 3558 DeepRsq: an improved imputation quality metric via a machine learning method
PrgmNr 3559 ExPheWas: a browser for gene-based pheWAS associations
PrgmNr 3560 Gene-based association approach in family samples using GWAS summary statistics
PrgmNr 3561 Gene-burden tests, gene-environment interactions and time-to-event data analysis within an efficient whole genome regression framework for large-scale biobanks
PrgmNr 3562 Gene-environment interaction testing using machine learning approaches and robust test statistics: a sex-specific risk score component for lung function
PrgmNr 3563 HaploSoup: IBD-based encoding for improved PCA genome matching
PrgmNr 3564 High-dimensional regression of continuous secondary traits under extreme phenotype sampling
PrgmNr 3565 Human Genetics Structure Activity Relationship (HG-SAR): A novel methodology to link human genotype-phenotype associations to high throughput functional activity data and validate targets for human disease
PrgmNr 3566 Identity-by-descent mapping in biobank-scale datasets
PrgmNr 3567 Individuals with limited healthcare interactions may degrade polygenic risk score analyses
PrgmNr 3568 Interactive Visualization of Relatedness and Ancestry Inference
PrgmNr 3569 Interpretable prioritization of causal genes in unsolved GWAS loci
PrgmNr 3570 LD Score regression (LDSC) with more comprehensive catalogs of LD scores
PrgmNr 3571 Leveraging functional annotation in a fast and powerful eQTL weighted gene-based association test
PrgmNr 3572 Leveraging Health Systems Data to Characterize a Large Effect Variant Conferring Risk for Liver Disease in Puerto Ricans
PrgmNr 3573 Leveraging simulations in extended pedigrees for population genetic analysis
PrgmNr 3574 MAST-Decon: Smooth Cell-type Deconvolution Method for Spatial Transcriptomics Data
PrgmNr 3575 Methods for genetic association analysis of mother-child pairs
PrgmNr 3576 Multiethnic joint analysis of summary statistics from genome-wide association studies
PrgmNr 3577 Multi-trait analysis of gene-environment-wide association using MAGENTA
PrgmNr 3578 Multivariate modeling of direct and proxy GWAS indicates substantial common variant heritability of Alzheimer's disease
PrgmNr 3579 Novel Global Quantile Regression Based Method for Differential Gene Expression Analysis of Quantitative traits
PrgmNr 3581 Reconstructing SNP Allele and Genotype Frequencies from GWAS Summary Statistics
PrgmNr 3582 SparsePro: an ultra-efficient Bayesian fine-mapping method integrating summary statistics and functional annotations
PrgmNr 3583 The power of pathway-specific polygenic risk scores

PrgmNr 3585 Using metagenomics to assess global levels of antimicrobial resistance

PrgmNr 3586 Utility of admixed individuals in genome wide association studies: A test case in COVID-19

PrgmNr 3587 Utilizing phenotype risk scores and genetically-regulated expression in a clinical biobank to investigate the genetic architecture of Mendelian disease genes

PrgmNr 3588 A phenome-wide Mendelian randomization study of plasma triglycerides and more than 6,000 diseases and traits

PrgmNr 3589 Assessing the difference in multiple sclerosis age of onset amongst different racial populations through the use of electronic health records

PrgmNr 3590 Bayesian analyses of variant-specific features enable penetrance estimates across cardiac channelopathies

PrgmNr 3591 BrainXcan identifies brain features associated with Alzheimer's disease using large scale genetic and neuroimaging data

PrgmNr 3592 Circulating proteins to predict adverse COVID-19 outcomes

PrgmNr 3593 Detecting DHCR7 variants consistent with Smith-Lemli-Opitz syndrome in autism patients

PrgmNr 3594 Dissecting the causal relationships between obesity and multiple sclerosis risk

PrgmNr 3595 Evaluation of genetic risk scores for incisional surgeries in primary open-angle glaucoma patients

PrgmNr 3596 Genetic Variation in Genes Associated with GERD in COPD

PrgmNr 3597 Genome-wide survival analysis of dental caries incidence

PrgmNr 3598 GWAS of self-report and imputed phenotypes capture distinct but complementary genetic architecture of stuttering

PrgmNr 3599 Identification of Genetic Loci Associated with Cognitive Preservation in the Midwestern Amish

PrgmNr 3600 Investigating White Blood Cell Count and Platelet Biology as a Possible Driver of the Comorbidity Between Coronary Artery Disease and Major Depressive Disorder

PrgmNr 3601 Joint intron splicing-based transcriptome-wide association study identifies new candidate susceptibility genes for breast cancer

PrgmNr 3602 Overwhelming numbers of patients who carry pathogenic familial hypercholesterolemia variants are predicted to have low clinical risk in the UK Biobank

PrgmNr 3603 Polygenic risk score analysis provides evidence of unique genetic architecture of Alzheimer’s disease in an Amish population

PrgmNr 3604 Prioritizing causal risk factors for severe COVID-19: an exhaustive Mendelian randomization study

PrgmNr 3605 Rare Variants in Non-coding Regulatory Regions and Non-syndromic Orofacial clefts

PrgmNr 3606 Sex-specific analyses in inflammatory bowel disease in African Americans reveal new loci on chromosomes 4, 5, 6, and 13

PrgmNr 3607 The Smoking Independent Causal Effects of Alcohol and Socioeconomic Traits on Lung Cancer Development

PrgmNr 3608 Today’s genetic risk scores for primary open-angle glaucoma underperform in African-descent participants from the Veteran Administration’s Million Veteran Program (MVP)

PrgmNr 3609 Transcriptome-wide association studies identify genetic determinants of brain-imaging phenotypes

PrgmNr 3611 Utilizing PrediXcan to detect Crohn's disease susceptibility genes in African Americans

PrgmNr 3612 A Translocation t(12;22) in Myxoid/Round Cell Liposarcoma

PrgmNr 3613 Amplification of CCND1 in Invasive High Grade Urothelial Carcinoma

PrgmNr 3614 Amplification of NMYC in Neuroblastoma

PrgmNr 3615 An amplification of EGFR (7p12) in a patient with Glioblastoma

PrgmNr 3616 An automated classification system for copy number variants based on 2019 ACMG guidelines

PrgmNr 3617 Characterizing tumor-infiltrated immune cells with spatial context using RNAscope-immunohistochemistry co-detection workflow in FFPE tissues

PrgmNr 3618 Monosomy 21, a rare finding in an elder patient with non-CLL-like monoclonal B-cell lymphocytosis

PrgmNr 3620 Outcomes from integrating an accessible delivery model for hereditary cancer risk assessment and genetic testing in populations with barriers to access

PrgmNr 3621 Patients with Birt-Hogg-Dubé syndrome and papillary renal cell carcinoma reveal a rare intragenic deletion in FLCN

PrgmNr 3622 Rapid whole genome sequencing (rWGS) on a novel digital microfluidic system: “on demand” personal automation for diagnosis of critically ill infants

PrgmNr 3623 Simplified and robust library construction for high-throughput HiFi sequencing for human variant detection

PrgmNr 3624 Utilizing Somatic Tumor Profiling To Identify Candidate Patients For Germline Genetic Testing In A Community Setting Reveals Barriers To Referral To Genetic Counseling

PrgmNr 3625 Verification of an EZH2 mutation test from formalin-fixed paraffin-embedded (FFPE) follicular lymphoma tumor tissue specimens

PrgmNr 3626 A large genome-wide association study for melanoma-specific survival reveals two novel loci linked to PSEN2, and ADCK3 genes

PrgmNr 3627 A segregation analysis of 17,425 population-based breast cancer families: Implications for breast cancer genetic susceptibility and risk prediction

PrgmNr 3628 Associations between tumor inflammation, survival, and somatic mutations in colorectal cancer patients

PrgmNr 3629 Associations of genetically predicted blood protein biomarkers with prostate cancer risk: a study using comprehensive protein genetic prediction modes

PrgmNr 3630 Breast cancer polygenic risk scores and rare variants in Latinas

PrgmNr 3631 Genetic risk of second primary cancer in breast cancer survivors: the Multiethnic Cohort Study

PrgmNr 3632 Improving risk prediction: Validation and exploration of breast cancer PRSs in a Midwest-American population

PrgmNr 3633 Increased burden of pathogenic variants in cancer predisposition genes in minority Americans with a cancer diagnosis between age 0-26: Use of linked California health registries

PrgmNr 3634 Inherited TP53 variants are associated with an increased risk of prostate cancer

PrgmNr 3635 The immuno-genetics of viral antigen response influence glioma susceptibility and survival in a subtype-specific manner

PrgmNr 3636 CDH1 and CTNNA1 variants with incomplete penetrance in Hereditary Diffuse Gastric Cancer Chilean families

PrgmNr 3637 Saturation Genome Editing of PALB2 Reveals Functionally Abnormal Missense Variants

PrgmNr 3638 A scalable deep learning framework for breast cancer prediction using DNA methylation data

PrgmNr 3639 Cell-type specific methylation changes from bulk whole genome bisulfite sequencing of matched primary and recurrent high grade serous ovarian cancers indicates stability of methylation landscape through tumor progression and chemoresistance

PrgmNr 3640 Exome-wide rare variant association study in lymphoid cancers

PrgmNr 3641 Identifying non-coding drivers of ovarian cancer by converging germline variants and somatic mutations

PrgmNr 3643 Spatial molecular profiling of Triple-Positive Ductal Carcinoma In Situ of the Breast using Visium Spatial Gene Expression for FFPE

PrgmNr 3644 Tracing intratumoral genetic heterogeneity using genetic barcoding in small cell lung cancer

PrgmNr 3645 Transcriptome-wide association study identifies novel genes associated with the immune traits in cancer

PrgmNr 3646 Epigenome-wide association study (EWAS) for soluble CD14 level in African Americans from the Jackson Heart Study

PrgmNr 3647 Exploring the genetic architecture of spontaneous coronary artery dissection using whole genome sequencing

PrgmNr 3648 Gene expression analysis of lipid traits in multi-ethnic population

PrgmNr 3649 Investigating a potential causal relationship between maternal blood pressure during pregnancy and future offspring cardiometabolic health

PrgmNr 3650 Investigating exposure mediated genetic scoring for coronary artery disease

PrgmNr 3651 Investigation of eQTLs in the CELSR2/ PSRC1/SORT1 gene cluster for independent effects in the liver and skeletal muscle that influence LDL levels and energy expenditure in American Indians

PrgmNr 3652 Methylome-wide association analysis identified novel CpG sites for lipid concentrations in ancestral-diverse populations from the PAGE Study

PrgmNr 3653 Polygenic risk score analysis of congenital heart disease phenotypes: elucidating the contribution of common variants to congenital heart defects

PrgmNr 3654 Response to Polygenic Risk: Results of the MyGeneRank Mobile Application-Based Coronary Artery Disease Study

PrgmNr 3655 Trans-ancestry PRSs of lipid traits to predict cardiovascular disease in the ancestrally diverse PAGE study

PrgmNr 3656 Biometric trait sex heterogenous SNPs disproportionately influence predicted gene expression and diverse traits later in life

PrgmNr 3657 Clinically impactful non-glycemic genetic effects on HbA1c yield missed detections of pre-diabetes, especially in African ancestry

PrgmNr 3658 Genome-wide association study of sphingolipid traits in diabetic patients reveals distinct cis-regulatory regions for sphingolipid pathway-associated genes

PrgmNr 3659 Grip strength, adiposity, and genetic risk of non-alcoholic fatty liver disease: Findings from the UK Biobank study

PrgmNr 3660 Integration of transcriptome genetics, metabolic perturbations, and trait associations identifies high-confidence cardiometabolic risk genes

PrgmNr 3661 Investigation of Rapid-onset Obesity with Hypothalamic dysfunction, Hypoventilation and Autonomic Dysregulation (ROHHAD) by comprehensive whole genome sequencing

PrgmNr 3662 Trans-ancestry meta-analysis of the X chromosome for obesity-related traits in over 600K individuals

PrgmNr 3663 Ancestry-aware polygenic risk scores identify host factors behind susceptibility and disease outcomes associated to SARS-CoV-2 infection from a nationwide longitudinal direct-to-consumer survey of 30,000 individuals

PrgmNr 3665 Common and rare variant analyses combined with single-cell multiomics reveal cell-type-specific molecular mechanisms of COVID-19 severity

PrgmNr 3666 Genome-wide association analyses of common infections in a large practice-based biobank

PrgmNr 3667 Mid-density Gene Expression Profiling of SARS-COV-2 infected samples using Applied Biosystems TaqMan Flexible Array Panels

PrgmNr 3668 Multi-ancestry GWAS reveals association between absolute neutrophil count and SETD1B

PrgmNr 3669 Predicting host response to Covid-19

PrgmNr 3670 Identification of Gene Expression Patterns Associated With Polycystic Ovary Syndrome

PrgmNr 3671 Mendelian randomization analyses reveal no evidence of causal relationships between chronic health conditions and sporadic and recurrent miscarriage

PrgmNr 3672 Development of a discovery pipeline for structural variation contributing to the cause of amyotrophic lateral sclerosis

PrgmNr 3673 First joint exome and metabolome analysis in dyslexia implicates immune system deficits and dysregulated sensory perception

PrgmNr 3674 Genome Wide Association Study of QSM brain measures in UK biobank to undercover the genetic basis of brain iron load

PrgmNr 3675 Integrated gene analyses of de novo mutations from 46,612 trios with autism and developmental delay

PrgmNr 3676 Multivariate genomewide associations on microstructural compartments of human brain highlight the consistent neurodevelopmental molecular pathways from adolescent to late adulthood

PrgmNr 3677 Pairwise contributions of 16p11.2 CNV genes to brain-related traits

PrgmNr 3678 Association between complement component 4 variation and childhood brain and behavioral phenotypes: Multiethnic PheWAS in the Adolescent Brain and Cognitive Development (ABCD) Study

PrgmNr 3679 Polygenic risk and causal inference of psychiatric comorbidity in inflammatory bowel disease

PrgmNr 3680 Polygenic risk scores for Bipolar Disorder predict age of diagnosis and the number of lifetime psychiatric diagnoses experienced by participants in the Australian Genetics of Bipolar Disorder Study

PrgmNr 3681 A GWAS of atopic dermatitis in African American individuals

PrgmNr 3683 A repetitive request: for the genome wide study of short tandem repeats in human phenotypic variation

PrgmNr 3684 Enabling biobank-scale GWAS data sharing with efficient compression and indexing

PrgmNr 3685 Genetic burden of cardio-metabolic diseases on women's health outcomes from a study of multi-ancestry population

PrgmNr 3686 Genetic variants affecting the expression of two Mendelian deafness genes are associated with age-related hearing impairment

PrgmNr 3687 Low-Coverage Whole Genome Sequence Genotyping Pipeline on Heterogeneous Stock Rats

PrgmNr 3689 Accelerated epigenetic aging in newborns with Down syndrome

PrgmNr 3690 Characterizing molecular mechanisms of the oncogene transcription factor FOXM1 in lung and breast cancer to suppress cell proliferation

PrgmNr 3691 Comparative analysis of bulk RNA versus single-cell RNA sequencing to profile gene expression in cellular subgroups within peripheral blood mononuclear cells

PrgmNr 3692 Genome-wide effects of sex chromosome aneuploidy on gene expression and DNA methylation in isogenic neural precursor lines

PrgmNr 3693 Interrogation of allergy associated regulatory variants for GATA3 in human T cells

PrgmNr 3694 Novel insight into the etiology of ischemic stroke gained by integrating human transcriptome wide association study with rodent expression data

PrgmNr 3695 Prediction of genetically regulated expression of asthma target tissues for African-ancestry populations

PrgmNr 3696 Regulatory landscape of CD4+ T cells with implications for asthma

PrgmNr 3697 Single-cell RNA sequencing of lung identifies disease-associated cell type-specific eQTLs

PrgmNr 3698 Association between the tobacco-specific carcinogen metabolite NNAL and DNA methylation in a multiethnic population

PrgmNr 3699 Automated High-throughput IP Workflows using Covaris Adaptive Focused Acoustics (AFA) System for Accelerated Epigenomic Mapping

PrgmNr 3700 Black Representation in Genomics Research (BRGR) Study of eQTLs in venous blood of 23andMe research participants with African ancestry

PrgmNr 3701 Circular RNAs preferentially localize to synapses in the human brain implying critical functions for synaptic plasticity and may be important components of neurodegeneration

PrgmNr 3702 Dissecting the evolutionary and functional architecture of human enhancer sequences with multiple ancestral origins

PrgmNr 3703 Enhancer enhancer interaction networks involved in long-distance genome regulation link multiple non-coding variants to function

PrgmNr 3704 Human-specific structural variants alter cis-regulation and chromatin structure

PrgmNr 3705 Integrative single cell analysis of cardiogenesis identifies developmental cellular trajectories and prioritizes noncoding mutations in congenital heart disease
PrgmNr 3706 Investigating conservation of chromatin accessibility signatures during iPSC reprogramming across human and mouse ATAC-seq data
PrgmNr 3707 Systematic reconstruction of the cellular trajectories of mammalian embryogenesis
PrgmNr 3708 Telomere-to-telomere map of a human epigenome
PrgmNr 3709 The landscape of alternative polyadenylation in single cells of the developing mouse embryo

PrgmNr 3710 Universal DNA methylation age across mammalian tissues
PrgmNr 3711 Using Taqman assays to analyze eQTL loci arising from GWAS Studies
PrgmNr 3712 A statistical framework for single-molecule transcription factor footprinting
PrgmNr 3713 Analyzing the Effects of GWAS Fine-Mapping on PGS Analyses for Type 2 Diabetes
PrgmNr 3714 ChromBPNet: Deep learning models of base-resolution chromatin profiles reveal cis-regulatory syntax and regulatory variation
PrgmNr 3715 Deep Learning Modeling of Transcription Factor Binding Specificity Using DNA Biophysical Properties
PrgmNr 3716 DMRscaler: A scale-aware method to identify regions of differential DNA methylation spanning basepair to multi megabase features
PrgmNr 3717 Outlier gene detection to improve biological validation and classification performance in RNA sequencing
PrgmNr 3718 A Genealogical Estimate of Genetic Relationships to Improve Detection of Population Structure Over Time
PrgmNr 3719 Causal variant effect sizes of complex traits differ between populations
PrgmNr 3720 Codon bias analysis of wildtype SARS-CoV-2 and prevalent variants indicates a mutational trajectory independent of human pulmonary tissue and genomic bias
PrgmNr 3721 Coevolution is pervasive between unrelated glycosylation pathways and points to potential disease modifiers
PrgmNr 3722 DNA methylation patterns underlying lifespan differences in mammals
PrgmNr 3723 Gnomix: local ancestry inference for whole genome sequences and large datasets
PrgmNr 3724 Human-specific evolution at multiple scales and dimensions
PrgmNr 3725 Identification of ancestry specific health risks in a large cosmopolitan biobank
PrgmNr 3726 Inferring negative natural selection at short tandem repeats in the human genome
PrgmNr 3727 Inferring population structure in biobank-scale genomic data
PrgmNr 3728 Population level study of tandem repeats using an ensemble call-set
PrgmNr 3729 Reconstructing spatio-temporal patterns of admixture in human history using present-day and ancient genomes
PrgmNr 3730 Reconstructing the 3D chromatin organization of archaic hominins reveals that genome folding shaped human-Neanderthal phenotypic divergence and introgression
PrgmNr 3731 The mutational dynamics of short tandem repeats in large, multigenerational families
PrgmNr 3732 The waiting distance distribution in ancestral recombination graphs
PrgmNr 3733 A retrospective study of health issues being discussed in a prenatal setting with female carriers of fragile X syndrome
PrgmNr 3734 Analysis of Financial Barriers Experienced by Prospective Genetic Counseling Students
PrgmNr 3735 Attitudes towards polygenic risk testing for glaucoma in an Australian population
PrgmNr 3736 Authorship Guidelines and Implementation: Promoting Equity and Inclusion
PrgmNr 3737 Clinical Sequencing Evidence-Generating Research (CSER) consortium Diversity and Inclusion Statement: development, revision, and adoption
PrgmNr 3738 Cloud-based biomedical data storage and analysis for genomic research: landscape analysis of data governance in emerging NIH-supported platforms

PrgmNr 3739 Conducting clinical genomic medicine research during the COVID-19 pandemic: Lessons learned from the CSER consortium experience

PrgmNr 3741 Designing and launching a pilot race and genetics seminar series for public health genetics students

PrgmNr 3742 Ethnic Differences in the Frequency of Cancer Reported from Family Pedigrees in the Prenatal Genetic Counseling Setting

PrgmNr 3743 Exploring a role for genetic counselors in changing client health behavior: An Australasian focus group study

PrgmNr 3744 Impact of the COVID-19 pandemic on the training and well-being of Medical Genetics and Genomics trainees

PrgmNr 3745 Medical genetics among Internal Medicine residents

PrgmNr 3746 Perspectives of Genetics Professionals on the Reporting of Variants of Uncertain Significance (VUS): Should they always be returned?

PrgmNr 3747 Population genetic screening: What are the psychosocial impacts?

PrgmNr 3748 Preparing the next generation of healthcare providers for precision medicine in practice: A social justice and equity-oriented clinical genomics training program for pre-clerkship medical school students

PrgmNr 3749 Reframing translational research in rare diseases in the United States: A mosaic approach

PrgmNr 3750 Researcher Views on Multi-omics Return of Results in the Molecular Transducers of Physical Activity Consortium (MoTrPAC)

PrgmNr 3751 Risk-reducing surgery in unaffected individuals receiving cancer genetic testing for hereditary breast and ovarian cancer in an integrated health care system

PrgmNr 3752 SeqScreener: a cloud-based software application for analyzing CRISPR experiments with Sanger Sequencing data

PrgmNr 3753 Splice-Switching Oligonucleotide-mediated Correction of a Deep Intronic Splice-variant in TIMMDC1 in Cells of Patients with Severe Early Onset Neurodegenerative Disorder

PrgmNr 3754 An Embedded ELSI Framework for the Human Pangenome Reference Consortium

PrgmNr 3755 Determining Clinical Relevance of Genetic Variants in Newborn Screened Conditions

PrgmNr 3756 Experience from the 3D health pilot study at UCSF: how to implement and the utility of large scale genomic testing for healthy adult population

PrgmNr 3757 Functional annotation of human mitotic gene domains using CRISPR/Cas9 tiling screens

PrgmNr 3758 IEMbase: An online knowledge base and mini-expert tool for the diagnosis of inborn errors of metabolism

PrgmNr 3759 Rarediseasegenes.com/phex: A comprehensive locus specific database of PHEX gene variants associated with X-linked hypophosphatemia vastly increases the number of known variants

PrgmNr 3760 Re-architecture of ClinGen's Variant Curation Interface (VCI) to Support Future Enhancements Aimed at Scaling Variant Classification

PrgmNr 3761 Significant Sparse Polygenic Risk Scores in 428 traits in UK Biobank

PrgmNr 3762 Specifications of the ACMG/AMP variant curation guidelines for MYOC: Recommendations from the ClinGen Glaucoma Expert Panel

PrgmNr 3763 The Atlas of Variant Effects (AVE) Alliance:understanding genetic variation at nucleotide resolution

PrgmNr 3764 Virtual Registry of 3,500 Leigh Syndrome and Primary Mitochondrial Disease Cases Constructed through Expert Curation and Semi-automated Literature Mining at MSeqDR

PrgmNr 3765 A rare case of Bardet Biedl syndrome presented as acute onset of heart failure and dilated cardiomyopathy

PrgmNr 3766 Further delineating Baraitser Winter syndrome through ACTB and ACTG1 phenotypic characterization

PrgmNr 3767 Lessons learned from the first 170 trios from the Genomic Autopsy Study

PrgmNr 3768 Parental Mosaicism in "De Novo" Tuberous Sclerosis Complex

PrgmNr 3769 De novo heterozygous POLR2A variant associates with autism spectrum disorder with epilepsy, hypotonia, strabismus and self-injurious behaviours

PrgmNr 3770 A CRISPR screen for modifiers of the rare disease DPAGT1-CDG (CDG-Ij)

PrgmNr 3771 A novel non-canonical splice site variant in Delta 4-Desaturase, Sphingolipid 1 (DEGS1) leads to loss of normal DEGS1 splicing and results in functional deficiency of ceramide desaturase activity in two patients with hypomyelinating leukodystrophy

PrgmNr 3772 An in vivo repurposing screen identifies novel therapeutic candidates for NGLY1 deficiency

PrgmNr 3773 Epilepsy gene variants increase risk of seizure onset in familial Cerebral Cavernous Malformation

PrgmNr 3774 Genetic modifiers of NGLY1 deficiency identified through a Drosophila genetic screen point to the role of NGLY1 in ERAD

PrgmNr 3775 Human genetics is great: Bone Mineral Density Studies in Patients with IBMPFD using DEXA

PrgmNr 3776 Identification of small and large variants in familial amyotrophic lateral sclerosis

PrgmNr 3777 Identification of X-linked missense variants in TAF1 in 4 unrelated families with autism spectrum disorder (ASD)

PrgmNr 3778 Loss of function variants in DIP2C are associated with neurocognitive delays

PrgmNr 3779 A mouse model recapitulating Bruck Syndrome provides insight into the role of Plod2 in bone and cartilage

PrgmNr 3780 Expanding clinical and molecular evidence of CYP26B1 involvement in retinoic acid catabolism defects

PrgmNr 3781 FDXR Mitochondriopathy: The challenges of reclassifying variants of unknown significance in a newly described syndrome

PrgmNr 3782 Long-Chain Fatty Acid Oxidation Disorder Gene Variants Identified Through a Gene Panel Sponsored Program from Clinically Confirmed and Suspected LC-FAOD

PrgmNr 3783 Novel heterozygous OPA3 mutation in a family with congenital cataracts, sensorineural hearing loss and neuropathy, without optic atrophy

PrgmNr 3785 Recessive genetic risk factor for hemochromatosis is associated with damaging iron deposition in motor circuits of the human brain

PrgmNr 3786 RNA-sequencing of tissue from family members identifies causal mutation in children

PrgmNr 3787 Transcriptomic profiling of paired skeletal muscles from healthy donors using bulk RNAseq and snRNAseq to identify DMD susceptibility factors

PrgmNr 3788 Translating precision medicine into the world of speech-language pathology: First follow-up results of the proactive Babble Boot Camp© intervention in infants with classic galactosemia

PrgmNr 3789 A Clinicopathologic Study of Malignancy in VCP Associated Multisystem Proteinopathy

PrgmNr 3790 Childhood and early onset glaucoma classification and genetic profile in a large Australasian disease registry

PrgmNr 3791 Deciphering the role of PTPN14 in BMP9 signaling and vascular malformations in Hereditary Hemorrhagic Telangiectasia

PrgmNr 3792 Genotype-phenotype correlations within type IV collagen glomerulopathy

PrgmNr 3793 Lower penetrance estimates for neurosusceptibility loci using an improved formula

PrgmNr 3795 Analytical performance of germline genome sequencing for clinical exome diagnostic testing

PrgmNr 3796 Clinical validation of germline genome sequencing copy number variant confirmation using qPCR

PrgmNr 3797 Detecting absence of heterozygosity using high-resolution optical genome mapping

PrgmNr 3798 Detection of RHD/RHCE hybrid genes using the Axiom™ Universal Blood Donor Typing research array

PrgmNr 3799 Identification of structural variants in constitutional disorders by optical genome mapping

PrgmNr 3800 Predicting Genes from Phenotypes using Human Phenotype Ontology (HPO) terms

PrgmNr 3801 Resolving complex pathogenic alleles using HiFi sequencing for long-range amplicon data with a new clustering algorithm

PrgmNr 3803 Biallelic loss of function ATM due to pathogenic synonymous and novel deep intronic variant c.1803-270T>G identified by genome sequencing in child with ataxia-telangiectasia

PrgmNr 3804 End ALS Kaggle challenge: Insights from aberrant transcriptome analysis

PrgmNr 3805 Establishing a platform for the functional study of oligogenic novel ALS genes

PrgmNr 3806 High-throughput classification of pathogenic SLC6A1 variants

PrgmNr 3807 Immune cell specific epigenetic and regulatory mapping of variants associated with multiple sclerosis in hispanic populations

PrgmNr 3808 Paternal allele of origin, radiosensitivity and the effect of genetic modifiers on the 3q29 Microdeletion Syndrome phenotype

PrgmNr 3809 Rapid Identification of Genetic Factors Contributing to Autism Spectrum Development with Disproportionate Megalencephaly in a Model System

PrgmNr 3810 Regulating translation of the GJB1 5' UTR

PrgmNr 3811 Transcriptome analysis provides insights into the possible function(s) of SPATA5L1

PrgmNr 3813 Comprehensive survey of the functional impact of SNP accessible variants in PSAT1 using a high throughput yeast assay

PrgmNr 3814 Development of multiplex synthetic positive controls for an expanded CFTR mutation testing

PrgmNr 3815 Down syndrome in patients with otitis media is associated with changes in the nasopharyngeal microbiota

PrgmNr 3816 Dynamic cell-type-specific eQTL from single-cell RNA-seq of inflamed and non-inflamed colon elucidate biological mechanisms of ulcerative colitis

PrgmNr 3817 Functional Validation of SOX2 Variants Contributing to Isolated GnRH Deficiency

PrgmNr 3818 Harnessing natural genetic variation in Drosophila to characterize the underlying mechanisms of ER stress preconditioning reveals role of histone methylation

PrgmNr 3819 High throughput, yeast-based functional assays of genetic variants in human genes associated with inherited anemias

PrgmNr 3822 90% of cryptic-donors activated in genetic disorders are present in variant-free RNA-Seq samples

PrgmNr 3823 Automated annotation of human centromeres

PrgmNr 3824 Cell type deconvolution of whole blood bulk RNA-Seq to reveal biological insights

PrgmNr 3825 Comprehensive variant calling in the GBA gene using whole-genome sequencing data

PrgmNr 3826 Computational framework for consensus variant calling and variant call aggregation

PrgmNr 3827 Evaluation of genome similarity for a twin family using 23andMe genotyping technology

PrgmNr 3828 Findings from the Critical Assessment of Genome Interpretation, a community experiment to evaluate phenotype prediction

PrgmNr 3829 Full lifecycle continuous workflow-enabled variant benchmarking

PrgmNr 3830 Length Variability in Short Tandem Repeats and Its Applications

PrgmNr 3831 Maverick: Variant prioritization for Mendelian diseases using deep learning

PrgmNr 3832 Predictive Interpretation and Scoring Model (PrISM) increases efficiency of variant classification by reducing manual review time

PrgmNr 3833 Small indel detection within VNTRs using short-read sequencing data

PrgmNr 3834 Systematic evaluation and Improvements for trans-eQTL Detection Methods Allows Identification of Novel trans-eQTLs in the GTEx data

PrgmNr 3835 Towards evidence-based recommendations and protocols for the use of computational tools in missense variant pathogenicity classification

PrgmNr 3836 Variant Calling of 50,000 UK Biobank Exomes in a Weekend

PrgmNr 3837 VizCNV: A tool for orthogonal CNV validation using whole genome sequence data

PrgmNr 3838 Benefits of Hybrid Capture for Viral Surveillance of SARS-CoV-2

PrgmNr 3839 First Report of Spatial Whole Transcriptome Profiling of Histological Structures of Multiple Organs

PrgmNr 3840 Optical Genome Mapping for High Throughput Analysis of Repeat Expansion Disorders

PrgmNr 3841 Simultaneous isolation of high-quality RNA and DNA from post-mortem human central nervous system tissues for omics studies

PrgmNr 3842 Single cell RNA sequencing and binary hierarchical clustering identify distinct normoxia and hypoxia associated interstitial macrophage cell types in mice exposed to hypoxia

PrgmNr 3843 Spatial Whole Transcriptome Analysis of human kidney histological Structures

PrgmNr 3844 Spatial whole transcriptome profiling uncovers unique functional insights into the histological structures of the human pancreas

PrgmNr 3845 Targeting clinically significant dark regions of the human genome with high-accuracy, long-read sequencing

PrgmNr 3846 Automated End-to-End NGS Sample Preparation: DNA Extraction and Library Preparation on a Novel Digital Microfluidics System

PrgmNr 3847 Comparison of fully automated to Semi-Automated Variant Curation in Genetic Carrier Screening

PrgmNr 3848 Designing Probesets for Interfering Variant Awareness on Genotyping Microarrays

PrgmNr 3849 DNAscopeTM: A novel chromogenic in-situ hybridization technology for high-resolution detection of DNA copy number and structural variations

PrgmNr 3850 Immunophenotyping of TCR and BCR clonotypes

PrgmNr 3851 Implications of biological networks: Search for the basal gene network using plasma proteomic signatures of COVID-19 patients

PrgmNr 3852 Large-scale functional assays to comprehensively assess SNP-accessible variants in the Urea Cycle Disorder genes, OTC, ASS1, and ASL

PrgmNr 3853 Optimizing performance of copy number variant calling based on improved baseline selection

PrgmNr 3854 Plasma protein profiling of Alzheimer’s and mild cognitive impairment subjects with a novel approach for identification of known and unknown candidate biomarkers

PrgmNr 3855 Proteogenomic analyses of non-small cell lung cancer subjects identifies candidate lung-cancer associated protein isoforms and protein variants

PrgmNr 3856 ProteographTM Analysis Suite: A cloud-scalable software suite for proteogenomics data analysis and visualization

PrgmNr 3857 Quantification of gene expression changes in mouse disease models using a high-throughput spatial omics platform

PrgmNr 3858 Reducing variability of breast cancer subtype predictors by grounding deep learning models in prior knowledge

PrgmNr 3859 "Drugging" PTPRD (AD, RLS and Addiction-associated gene) PTPRD

PrgmNr 3860 Development and optimization of a 43-gene pharmacogenomic panel using enrichment-based capture and PacBio HiFi sequencing

PrgmNr 3861 Genomic analysis of angiotensin converting enzyme (ACE) inhibitor - induced angioedema

PrgmNr 3862 Identification of pharmacogenetic predictors of methotrexate-induced mucositis in children with cancer

PrgmNr 3863 Latino and Indigenous American populations may have five times as many pharmacogenetic variants as Europeans yet remain underrepresented in pharmacogenetic studies

PrgmNr 3864 Research clinical pilot project for in silico pharmacogenomics concurrent with whole-genome sequencing in the pediatric ICU

PrgmNr 3865 Space-optimized HLA typing using Axiom™ PGx research solutions

PrgmNr 3866 The Study of HLA-B*48:01 and HLA-B*55:01 Frequency in Healthy Thai Population

PrgmNr 3867 A first exploration of the genetic architecture of linea nigra and its relationship with pigmentation

PrgmNr 3868 Clinical experience of an alpha thalassemia carrier screening assay with an increased detection rate due to novel variant calling

PrgmNr 3869 Developmental and temporal characteristics of clonal sperm mosaicism

PrgmNr 3870 Exploring the use of whole genome sequencing for carrier screening

PrgmNr 3871 Genematching candidates from the Genomic Autopsy Study to elevate variant classification for prenatal clinical use

PrgmNr 3872 Leveraging variant classification evidence from diagnostic genetic testing to improve the rate of positive results in reproductive carrier screening

PrgmNr 3873 Male microchimerism in females: A quantitative study of twin pedigrees to investigate mechanisms

PrgmNr 3874 Whole-exome sequencing confirms GDF15 is the greatest genetic risk factor for Hyperemesis Gravidarum and identifies rare and novel variants in appetite genes GDF15 and GFRAL

PrgmNr 3875 A variational Bayesian approach to characterize latent genetic components using GWAS summary statistics

PrgmNr 3876 Biobank-scale estimation of the proportion of trait variance explained by gene-environment interactions

PrgmNr 3877 Capturing fine-scale structure of single-cell expression and chromatin accessibility using a topic model

PrgmNr 3878 Combining SNP-to-gene linking strategies to pinpoint disease genes and assess disease omnigenicity

PrgmNr 3879 Development of a Polygenic Risk Score to Predict Severe Outcome of a COVID19 Infection

PrgmNr 3880 Estimating heritability explained by local ancestry and evaluating stratification bias in admixture mapping from summary statistics

PrgmNr 3881 Explainable and extendable machine learning models for identifying prognostic radiogenomic biomarkers from breast cancer multimodal imaging and genomic data

PrgmNr 3883 FDR confidence interval selection and adjustment approach for large-scale hypothesis testing

PrgmNr 3884 iCURL: Refining Estimated Identical-by-Descent Segments Via Community Detection

PrgmNr 3885 Improving Polygenic Risk Score Transferability by Leveraging GWAS Signals from Diverse Populations

PrgmNr 3887 Repeat Expansions and Somatic Instability Observation using Optical Mapping

PrgmNr 3888 Residual Proteome-wide Association Study Identifies Genes for Blood-Related Traits

PrgmNr 3889 Searching for consistent associations with a multi-environment knockoff filter

PrgmNr 3890 Shared Genomic Segment Analysis in a Large High-Risk Chronic Lymphocytic Leukemia Pedigree Implicates CXCR4 in Inherited Risk

PrgmNr 3891 The Dynamic and Heritable Epigenetic Landscape

PrgmNr 3892 Transethnic genetic-correlation estimation and partitioning

PrgmNr 3893 Using adopted singletons to partition maternal genetic effects into pre- and post-natal effects on offspring phenotype

PrgmNr 3894 A benchmarking study of SARS-CoV-2 whole-genome sequencing protocols using COVID-19 patient samples

PrgmNr 3895 An Epigenome-wide association study of body mass index (BMI) in the Multiethnic Cohort Study

PrgmNr 3896 Evidence of causality between type 2 diabetes and dementia in the Million Veteran Program

PrgmNr 3897 Examining genetic associations with liver steatosis in Mexican-origin adults

PrgmNr 3898 Genetic determinants of prostate-specific antigen levels improve cancer screening utility

PrgmNr 3899 Genetic factors explain varying glaucoma prevalence and pressure subtype across ancestries

PrgmNr 3900 Heritability of audiometric phenotypes using multigenerational pedigrees

PrgmNr 3901 Methylation patterns associated with inflammation traits in racially and ethnically diverse populations

PrgmNr 3902 ML-based COPD phenotype denoising using volumetric flow data improves genomic discovery and disease risk prediction

PrgmNr 3903 Multitrait and AI-enhanced GWAS of glaucoma and its risk factors enables PRS based prediction of risk which is portable across ancestries

PrgmNr 3904 Multi-trait Genome-wide association study identifies variants associated with corneal parameters that contributing to keratoconus risk

PrgmNr 3905 Perceived aging in the UK Biobank - A proxy measure for skin aging?

PrgmNr 3906 Quantifying uncertainty and variability of recessive disease prevalence using Monte Carlo estimates

PrgmNr 3907 Tissue-specific functional annotations highlight association of liver polygenic risk score with Alzheimer's disease and related biomarkers

PrgmNr 3908 Utility of polygenic risk scores for colorectal cancer risk assessment across diverse populations

PrgmNr 3909 mTOR inhibition improves lymphatic dysplasia caused byPIK3CAp.H1047L in preclinical models

Plenary Sessions

PrgmNr 1052 - Trans-ancestry meta-analysis enhances discovery, fine mapping and polygenic prediction of body mass index in continental Africa

View session detail

Author Block: **T. Chikowore**[1], M. Chong[2], L. K. Micklesfield[1], Z. Lombard[3], J. H. Goedecke[4], S. Fatumo[5], S. A. Norris[1], M. Ramsay[6], P. W. Franks[7], G. Pare[8], A. P. Morris[9]; [1]Univ. of the Witwatersrand, Johannesburg, South Africa, [2]Population Hlth.Res. Inst. (Genetic Molecular Epidemiology Lab), Hamilton, ON, Canada, [3]Natl. Hlth.Lab. Service, Johannesburg, South Africa, [4]South African Med. Res. Council, Capetown, South Africa, [5]London Sch. of Hygiene and Tropical Med., London, United Kingdom, [6]Univ. of the Witwatersrand, Johannesburg, Gauteng, South Africa, [7]Lund Univ., Malmo, Sweden, [8]Hamilton, ON, Canada, [9]The Univ. of Manchester, Manchester, United Kingdom
Disclosure Block: **T. Chikowore:** None.
Obesity is a major global public health problem, yet most genome-wide association studies (GWAS) for obesity traits have been performed in European ancestry populations. This lack of diversity impedes genetic discovery and limits the utility of genetics in obesity therapeutics. Here we report the first trans-ancestry meta-analysis for body mass index (BMI) involving continental Africans, and evaluate its utility in this population group. We aggregated GWAS of BMI in 678671 individuals from the United Kingdom Biobank (Europeans), Japan Biobank , African Partnership of Chronic Disease Research (APCDR), and the Population Architecture and Genetic Epidemiology (PAGE) who are representative of diverse global populations through inverse variance weighted, fixed-effects trans-ancestry meta-analysis implemented in Metasoft. We identified 660 loci at genome-wide significance (p -8), 164 of which have not been previously reported for BMI in any population. We compared the correlation of the effect sizes of lead variants from the trans-ancestry meta-analysis and those in the European GWAS in UK Biobank with ones found in a GWAS of 10900 continental Africans from the AWIGen study.For lead variants that were nominally significant in AWIGen, the effect size correlation of the trans-ancestry meta-analysis vs continental Africans was stronger (Pearson correlation = 0.45; p =0.0085) and significant compare to that with the European GWAS vs continental Africans (Pearson correlation = 0.35; p =0.084) . Trans-ancestry meta-analysis improved fine-mapping resolution over European ancestry participants in the UK Biobank as indicated by median 99% credible set sizes of 39 and 64, respectively. A polygenic risk score (PRS) derived from the trans-ancestry meta-analysis was more predictive of BMI in the continental Africans than the European ancestry-specific score. The mean difference in weight and BMI in the first and last deciles of the trans-ancestry PRS was 16.00kg and 6.45kg/m^2, respectively . The trans-ancestry PRS appeared to convey sex-specific effects (P_{int}=1.44 x10^{-66}) in continental Africans, such that the difference in BMI between first and last deciles was three-fold the magnitude in women (8.78kg/m^2; p =1.69 x10^{-92}) than men (2.90kg/m^2; p =4.86 x10^{-18}). Such an aspect has not been reported in Europeans.Our findings demonstrate that the inclusion of diverse populations in trans-ancestry meta-analyses improves fine-mapping resolution and enhance the transferability of PRS to African populations

PrgmNr 1053 - Investigating relative contributions to complex trait architecture from sequence elements originating across multiple evolutionary time-scales

View session detail

Author Block: **M. Sohail**[1,2], J. Novembre[3,4]; [1]Ctr. for Genomic Sci., Natl. Autonomous Univ. of Mexico (UNAM), Cuernavaca, Mexico, [2]Human Genetics, Univ. of Chicago, Chicago, IL, [3]Univ. of Chicago, Chicago, IL, [4]Dept. of Ecology and Evolution, Univ. of Chicago, Chicago, IL
Disclosure Block: **M. Sohail:** None.
Regulatory genomic elements have been found to be enriched for heritability across complex traits but can have different ages with respect to when they originated and became functionally constrained. We ask whether specific regulatory elements or sequences that have gained novel function during human evolutionary history are enriched in heritability for complex traits, and if this is enrichment is variable within and among trait domains. We partition heritability along the genome using S-LDSC and previously published genomic annotations from different periods of our evolutionary history. These are fetal and adult brain human-gained (HG) enhancers and promoters (H3K27as and H3K4me2 histone modification peaks in the cerebral cortex gained in humans compared to mouse and rhesus macaque), human accelerated regions, ancient selective sweeps, Neanderthal introgressed regions, and Neanderthal and Denisovan depleted regions. Meta-analyzing across 41 independent traits analyzed in Hujoel et al (2019) from UK Biobank and other GWAS, we find a significant heritability enrichment for fetal human gained enhancers and promoters expressed in the cerebral cortex at 7 post-conception weeks (pcw) and at 8.5 pcw, and those expressed in the occipital cerebral cortex (visual processing center) at 12 pcw (FDR

PrgmNr 1054 - Rare variant associations from exome sequencing of 454,787 individuals in the UK Biobank

View session detail

Author Block: **J. D. Backman**, A. H. Li, A. Marcketta, D. Sun, C. E. Gillies, D. Liu, A. E. Locke, S. Balasubramanian, A. Yadav, N. Banerjee, J. Kosmicki, J. Mbatchou, A. Damask, X. Bai, E. K. Maxwell, J. Mighty, Regeneron Genetics Center, M. B. Jones, L. J. Mitnaul, L. Habegger, W. Salerno, A. R. Shuldiner, L. A. Lotta, J. Overton, M. Cantor, J. G. Reid, H. M. Kang, J. L. Marchini, A. Baras, G. R. Abecasis, M. A. Ferreira; Regeneron Genetics Ctr., Tarrytown, NY
Disclosure Block: **J.D. Backman:** Salary/Employment; Regeneron Pharmaceuticals.
A major goal in human genetics is to use natural variation to understand the phenotypic and biological consequences of each of the protein coding genes in the genome. Towards that goal, we generated and analyzed exome sequencing of 454,787 participants of the UK Biobank (UKB) study, representing the largest catalog of protein-coding variation identified to date (~12 million variants, including ~8 million missense variants and ~1 million putative loss-of-function [pLOF] variants), exceeding the combined total of gnomAD and TOPmed reference datasets (~9 million variants). Using these data, we first evaluated genomic constraint at a resolution not previously possible in humans. For example, we identified exons depleted of nonsynonymous variation in genes such as *KCNQ2* and *U2AF2* (concordant with previous findings), suggesting that mutations in these highly constrained regions are likely to have deleterious phenotypic consequences. Conversely, we observed an enrichment of nonsynonymous variation in genes associated with clonal hematopoiesis of indeterminate potential (CHIP; eg. *DNMT3A*, *TET2*), most of somatic origin and consistent with positive selection.

We then assessed the impact of rare (MAFâ¤1%) pLOF and deleterious missense variants on 3,998 health- and behavior-related traits, testing the effect of variants individually and also on aggregate, through gene burden tests. We identified 8,865 significant associations with 492 traits at a Bonferroni-corrected significance threshold of Pâ¤$2x10^{-11}$. Key findings from these analyses include: 1) 564 genes associated with at least one trait; 2) 16% of associations identified with a burden of ultra-rare variants (MAFâ¤0.001%); 3) ~90% of associations remained significant after conditioning on nearby common variants identified through GWAS; 4) associations were highly concordant between European and other ancestries; 5) associations with rare variants were enriched in GWAS loci, and so can help prioritize genes underpinning common, non-coding variant associations; 6) significant associations with quantitative traits helped identify 13 protective sub-threshold (Pâ¤10^{-4}) associations with genetically-correlated diseases, such as between a burden of pLOF and deleterious missense variants in *SLC22A12*, lower serum urate, and protection from gout.

Overall, we show that the wealth of high-quality phenotype data available in the UKB study, in conjunction of whole-exome sequencing, provides an unprecedented opportunity to elucidate gene function through association studies of protein-coding variants.

PrgmNr 1055 - High-resolution structural and temporal mapping illuminates relationships between 3D chromatin structure, enhancer activity, and gene regulation during macrophage activation

View session detail

Author Block: **K. Reed**[1], E. S. Davis[1], D. Phanstiel[2]; [1]Univ. of North Carolina, Chapel Hill, Chapel Hill, NC, [2]UNC, Chapel Hill, NC
Disclosure Block: **K. Reed:** None.
DNA loops physically connect enhancers to the promoters of target genes, which can be over a million base pairs away. Many of these loops are cell-type specific, forming during differentiation and activation. While these condition-specific contacts are often accompanied by changes in gene expression and chromatin accessibility, the causal relationships are still unclear.
To address this gap, we have conducted a multi-omics time course to map structural and regulatory changes with high temporal resolution, focused on eight time points in the early stages of human macrophage activation. Integration of these datasets revealed the order of events, providing many insights into the relationships between genome architecture, enhancer activity, protein binding, and gene expression.
Deeply sequenced in situ Hi-C revealed significant rewiring of DNA loops during cellular activation, including some transient loops found only at intermediate time points, and others that changed as quickly as 30 minutes after treatment. RNA-seq, H3K27ac ChIP-seq and ATAC-seq identified massive changes in gene expression and chromatin accessibility.
In general, chromatin accessibility changed faster than most loops and genes. Among stable three-dimensional contacts, enhancers looped to genes showed significantly correlated patterns of activation when compared to un-looped matched pairs, with enhancer activation preceding gene activity by approximately 1 hour. Newly formed and strengthened loops correlated with increased internal and anchor gene expression. Surprisingly, we saw a striking anticorrelation between internal gene expression and contact frequency at loops that declined over the course of activation, suggesting that increased gene transcription may be disrupting loop extrusion in some circumstances.
On a coarser level, gene expression correlated with A/B compartmentalization as predicted, but changes in transcription appear to drive changes in compartmentalization to a lesser extent than expected. We found that a possible cause for this discrepancy is apparent poising of regions in the A compartment prior to gene activation in the region.
The integrated data from this study provides a deep resource for the continued exploration of structural and functional changes in the genome.

PrgmNr 1199 - Early detection of cancers using plasma cell-free DNA methylomes up to 7 years prior to clinical diagnosis

View session detail

Author Block: **N. Cheng**[1,2], D. Soave[3], K. Skead[4,2], T. Ouellette[1,2], S. Bratman[5,2], D. DeCarvalho[5,2], P. Awadalla[4,2]; [1]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Wilfrid Laurier Univ., Waterloo, ON, Canada, [4]Ontario Inst. Cancer Res., Toronto, ON, Canada, [5]Princess Margaret Cancer Ctr., Toronto, ON, Canada
Disclosure Block: **N. Cheng:** None.
Cancer survival rates are significantly improved when detected at early stages, particularly when the tumour is still localised to the tissue of origin. However, effective screening tools for early cancer detection is currently limited to a subset of cancer types. Profiling cell-free DNA (cfDNA) patterns has emerged as a prominent non-invasive biomarker for detection and subtyping of cancers. However, owing to difficulties in observing the early development of human malignancies as cancers are often detected once they become symptomatic, most cancer biomarker and evolution studies to date have primarily examined the genomics from solid tumour or liquid biopsies following a diagnosis. Utilizing cfDNA as a screening tool for early cancer detection requires profiling of blood plasma samples collected from asymptomatic individuals prior to the diagnosis of cancers to enable assessment of the earliest detectability and predictive performance of potential biomarkers. Here, we leverage the Canadian Partnership for Tomorrowâ□□s Health Project (CanPaTH), to profile blood plasma collected prior to the clinical detection of underlying cancers. Specifically, we use cell-free methylated DNA immunoprecipitation and high-throughput sequencing (cfMeDIP-seq), a highly sensitive assay for profiling cfDNA methylomes, to profile over 300 blood plasma samples collected up to 7 years prior to the detection of a breast, prostate or pancreatic cancer, in addition to matched controls with no history of cancer free through follow-up. We demonstrate using machine learning that not only can cfDNA methylation markers indicative of breast cancers be detected up to 7 years prior to clinical detection, predictive models classifying these individuals with an undetected cancer can achieve AUROC of 0.76 (95% CI: 0.72 - 0.81). We further highlight that cfDNA methylation markers in blood plasma collected close to diagnosis differ from those collected over 2 years prior to diagnosis revealing potential novel markers at the earliest stages of cancer development. In our current studies, we focus specifically on breast, prostate and pancreatic cancer cases, and are extending this to further pan-cancer applications in subsequent investigations.

PrgmNr 1200 - Genetic regulation of DNA methylation across tissues reveals thousands of molecular links to complex traits

View session detail

Author Block: **M. Oliva**[1], K. Demanelis[2], F. Jasmine[1], Y. Lu[1], GTEx Consortium, H. Hahsan[1], K. G. Muhammad[1], L. S. Chen[1], B. L. Pierce[1,3,4]; [1]Dept. of Publ. Hlth.Sci., Univ. of Chicago, Chicago, IL, [2]Dept. of Med., Univ. of Pittsburgh, Pittsburgh, PA, [3]Comprehensive Cancer Ctr., Univ. of Chicago, Chicago, IL, [4]Dept. of Human Genetics, Univ. of Chicago, Chicago, IL
Disclosure Block: **M. Oliva:** None.
Epigenetic modifications play a fundamental role in gene regulation in humans. **We generated DNA methylation (DNAm) data for 987 samples from the Genotype-Tissue Expression (GTEx) project, from 9 tissues** (breast, colon, kidney, lung, muscle, ovary, prostate, testis and whole blood) and 424 subjects. We mapped DNAm quantitative trait loci in *cis* (mQTLs), contrasted mQTLs with expression QTLs (eQTLs) features, and assessed their relative contributions to complex traits. We observed that while eQTL enrichment in transcription factors involved in basal transcription, **mQTLs tend to occur in binding sites of proteins involved in 3D organization of the genome**. Only 21% of mQTLs were identified as suggestively colocalized (PP4 > 0.5) with at least one eQTL. These results suggest that **mQTLs and eQTLs diverge in their underlying biological mechanisms** and that genetic co-regulation of DNAm and gene expression is not pervasive.
We integrated QTLs with genome-wide association study (GWAS) summary statistics of 87 GWASs and observed that **mQTL colocalizations are abundant and complementary to eQTL-GWAS colocalizations**: only 27% (749/2734) of GWAS hits colocalized with both QTL types, while 55% of hits colocalized with mQTLs exclusively. Among those, we identified an ovary-specific mQTL colocalized with a breast cancer GWAS signal in the known cancer-associated TERT locus, although TERT expression is generally undetectable in adult tissues. These results emphasize the importance of integrating multiple -omics to maximize the expectation of identifying molecular links to inheritable traits, and suggest that DNAm may be genetically co-regulated with gene expression in a particular context which causally impacts the trait; but unlike gene expression, only methylation is identifiable beyond the causal context. We observed that mQTL-GWAS colocalizations are enriched in pleiotropic variants regulating multiple CpG sites (mCpGs) (OR = 2.65, Fisherâ□□s exact test P 0.5) with ovarian cancer risk. These findings indicate that **mQTLs that exhibit regulatory pleiotropy have increased chances to impact a complex trait**, and that genetic variants can drastically modify the DNAm and gene expression landscape in a tissue-specific manner, impacting disease risk. The DNAm dataset generated herein enhances existing GTEx resources and enables the research community to investigate the inherited susceptibility to human disease from both cross-tissue and cross-omics perspectives.

PrgmNr 1201 - Widespread evidence of systematic bias in estimates of genetic correlation due to assortative mating

View session detail

Author Block: **R. Border**[1], G. Athanasiadis[2], A. Buil Demur[3], A. Schork[4], T. M. Werge[5], K. S. Kendler[6], J. Flint[7], A. Dahl[8], N. A. Zaitlen[9]; [1]UCLA, Los Angeles, CA, [2]Roskilde, Denmark, [3]Inst. of Biological Psychiatry, Roskilde, Denmark, [4]Res. Inst. for Biological Psychiatry, Roskilde, Denmark, [5]Univ. of Copenhagen, Roskilde, Denmark, [6]Virginia Commonwealth Univ, Richmond, VA, [7]WellcomeTrust Human Gen, Oxford, United Kingdom, [8]San Mateo, CA, [9]UCSF, San Francisco, CA
Disclosure Block: **R. Border:** None.
Positive genetic correlation estimates are commonly interpreted as evidence that two traits share overlapping biological underpinnings. With the increasing availability of large consortium and biobank datasets, considerable research activity has focused on cataloging the genetic correlation structure of complex human traits to better understand shared etiology and transdiagnostic risk factors. In the present work, we demonstrate that primary-phenotypic cross-trait assortative mating (AM) induces substantial genetic correlations among traits with independent genetic effects, which are then overestimated by widely-used marker-based estimators, including genomic-relatedness residual maximum likelihood and linkage disequilibrium score regression (LDSC). For example, a single generation of AM for two 50% heritable traits correlated across mates at r=0.5 will lead to LDSC genetic correlation estimates > 0.2 in the complete absence of pleiotropy. Using empirical cross-mate phenotypic correlation and heritability estimates across a broad array of previously studied anthropometric, metabolic, psychosocial, and psychiatric phenotypes measured in multiple large samples (n ranging between 44,686 and 500,000 individuals), we estimate the extent to which AM alone might plausibly account for previously published genetic correlation estimates. With respect to psychiatric disorders, we demonstrate that even small misdiagnosis errors together with cross-trait AM could account for previous genetic correlation estimates (i.e., without out any true pleiotropy) between some disorders (e.g., Schizophrenia and anxiety disorders), but not others (e.g., Schizophrenia and Bipolar Disorder). Averaging across disorders, AM alone would result in genetic correlation estimates 38.9% as large as those reported by previous studies. Finally, we provide evidence congruent with AM at the genetic level by examining cross-trait cross-chromosome polygenic score correlations. We discuss our results, which indicate that the genetic correlation is an unreliable indicator of shared etiology for the numerous trait-pairs subject to cross-trait AM, in the context of the widespread application of the genetic correlation as a direct measure of biological overlap or shared pathogenesis.

PrgmNr 1202 - Benefits and cost-effectiveness of cascade testing for pediatric cancer predisposition syndromes: Findings from the PreEMPT Model

View session detail

Author Block: **K. D. Christensen**[1,2], G. O'Brien[3], P. M. McMahon[1], N. K. Stout[1,2], J. M. Yeh[3,2], A. C. Wu[1,2,3]; [1]Harvard Pilgrim Hlth.Care Inst., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Boston Children's Hosp., Boston, MA

Disclosure Block: **K.D. Christensen:** None.

Background

Newborn population-based genetic screening may reduce mortality associated with pediatric cancers and could be cost-effective, especially as costs decline. Cascade testing of siblings of newborns with pathogenic variants could further improve population health.

Methods

The Precision Medicine Prevention and Treatment (PreEMPT) Model projects the incremental benefits, costs, and cost-effectiveness of newborn genetic screening. Using the model to estimate outcomes associated with a 11-gene panel of pediatric cancer predisposition syndromes (*RET, RB1, TP53, DICER1, SUFU, PTCH1, SMARCB1, WT1, APC, ALK,* or *PHOX2B*), we projected outcomes associated with cascade testing of siblings of newborns identified to be a pathogenic variant carrier via newborn screening. We assumed each newborn had one newborn sibling with a 50% chance of harboring the same variant. Benefits were modeled via reductions in advanced disease, cancer deaths, and treatment-related late mortality, assuming 100% adherence to surveillance.

Results

In a cohort of 3.7 million newborns, the model estimated 1,584 newborns had pathogenic variants and 792 siblings would have the same variants. An estimated 115 siblings with these variants would develop cancer before age 20 years. If these siblings underwent surveillance for early detection, 15 (95%UI, 11 to 20) cancer deaths would be averted, a reduction of 52% (95%UI, 45% to 59%). Compared with usual care, sibling cascade testing had an incremental cost-effectiveness ratio (ICER) of $17,950 per LY gained (95%UI, $7,070 to $31,450). Even if only 10% of mutation carriers identified in newborn screening had inherited mutations from parents (and therefore 5% of siblings were therefore expected to be carriers), the ICER increasing only to $23,800/LY gained (95% UI, $11,770 to $40,160).

Conclusion

Sibling cascade testing could potentially avert half of predicted cancer deaths and be a high-value intervention. Testing of siblings of newborns with pathogenic variants is likely an effective way to enhance cancer outcomes. Findings also highlight how targeted approaches can dramatically improve the value of genetic screening and be more reasonable strategies to achieve population-level benefits.

PrgmNr 1288 - Interplay between longitudinally-measured gene expression and metabolite levels in whole blood in the MultiMuTHER study

View session detail

Author Block: **J. El-Sayed Moustafa**[1], A. Ramisch[2,3], G. Leday[4], Y. Jiao[5,6,7], G. Nicholson[7], Y. N. Raza[1], M. Stevens[2,3], M. Abdalla[5,6,7], C. Menni[1], C. Holmes[7], T. D. Spector[1], M. I. McCarthy[5,6,8], S. Richardson[4], E. T. Dermitzakis[2,3], K. S. Small[1]; [1]Dept. of Twin Res. and Genetic Epidemiology, King's Coll. London, London, United Kingdom, [2]Dept. of Genetic Med. and Dev., Faculty of Med., Univ. of Geneva, Geneva, Switzerland, [3]Inst. of Genetics and Genomics in Geneva, Geneva, Switzerland, [4]Med. Res. Council Biostatistics Unit, Cambridge, United Kingdom, [5]Wellcome Trust Ctr. for Human Genetics, Nuffield Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, [6]Oxford Ctr. for Diabetes, Endocrinology and Metabolism, Radcliffe Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, [7]Dept. of Statistics, Univ. of Oxford, Oxford, United Kingdom, [8]Oxford NIHR BioMed. Res. Ctr., Oxford Univ. Hosp. Trust, Oxford, United Kingdom

Disclosure Block: **J. El-Sayed Moustafa:** None.

Multi-omic datasets represent a snapshot of the systemâ□□s physiological state, and are increasingly being considered for health monitoring and earlier identification of disease risk. Deeper understanding of the cross-sectional and longitudinal interplay between concurrently-measured â□□omics is critical to enabling a systems-wide approach to health monitoring.

We therefore explored univariate and multivariate cross-sectional and longitudinal patterns of variation in the MultiMuTHER study, which comprises gene expression (RNASeq - 16,292 genes) and metabolomics (Metabolon-1,197 metabolites) data in whole blood at three or more timepoints over up to eight years per individual from 335 TwinsUK subjects [age range 32-80; median=61yrs]. TwinsUK is a deeply-phenotyped cohort of twins with extensive â□□omics data and repeat phenotypic measures.

We assessed association between gene expression and metabolites, and identified 105,629 gene-metabolite associations significant at a 5% false discovery rate. Approximately 81% of genes and 99% of metabolites were found to have at least one significant gene-metabolite association. Expression of each gene was associated with a mean (1^{st}-3^{rd} quartiles) of 8 (2-10) metabolites, while each metabolite was associated with a mean (1^{st}-3^{rd} quartiles) of 116 (13-119) genes. Striking examples of hub metabolites and genes with extensive associations were identified, including nicotinamide, which was associated with expression of 5,081 genes, as well as a number of genes involved in fatty acid metabolism, including CPT1A, whose expression levels were associated with 309 metabolites, and ACAA2, whose expression levels were associated with 182 metabolites.

Longitudinal gene expression association analyses (GEAS) revealed expression levels of ~40% of 16,292 genes showed significant change over time (FDR 5%), while ~5% of 915 metabolites showed significant longitudinal variation (FDR 5%). Genes showing longitudinal change over time were found to have a higher number of gene-metabolite associations than those exhibiting stable expression over time (P -16), while metabolites exhibiting longitudinal variation did not.

In summary, we have performed one of the largest multi-omic longitudinal studies of concurrently-measured gene expression and metabolite levels in whole blood, identifying over 100,000 gene-metabolite associations. This study provides novel insight into the interplay between gene expression and metabolites, and may inform systems-wide approaches to projection of temporal progression of age-related diseases.

PrgmNr 1289 - Integrative single-cell eQTL analysis of 2.3 million cells elucidates causal gene expression mechanisms and relevant cellular contexts in 34 brain GWAS traits

View session detail

Author Block: **Y. Park**[1,2], L. He[3], N. Sun[4,5], H. Mathys[4,6], L. Hou[4,5], K. Galani[4,5], L-L. Ho[4,7], D. Bennett[8], L-H. Tsai[4,6], M. Kellis[4,5]; [1]Univ. of British Columbia, Vancouver, BC, Canada, [2]BC Cancer Res., Vancouver, BC, Canada, [3]Duke Univ., Durham, NC, [4]Massachusetts Inst. of Technology, Cambridge, MA, [5]Broad Inst. of MIT and Harvard, Cambridge, MA, [6]The Picower Inst. for Learning and Memory, Cambridge, MA, [7]Broad Inst. of MIT and Harvard, Canbrudge, MA, [8]Rush Univ., Chicago, IL
Disclosure Block: **Y. Park:** None.
Gene expressions in the human brain are maintained and orchestrated by regulatory-genomic models in multiple neuronal and glial cell types. Although common genetic variation may not immediately cause disease phenotypes, polygenic effects of cis-regulatory models of relevant cell types can accumulate over time and eventually alter the disease propensity.
To understand cell-type-specific regulatory mechanisms, we profiled single nucleus RNA-seq data of 2.3 million cells across 429 postmortem human brain samples, of which 396 individuals are genotyped by whole genome sequencing. After adjusting batch-specific effects, fitting a von Mises Fisher mixture model, we hierarchically annotated the cells into 16 distinctive cell types, which include (1) four cortical layers of excitatory neurons, (2) four types of inhibitory neurons, (3) four glial cell types, and (4) four vasculature cell types. In comparison with observed covariates, we found significant negative correlations between the fraction of somatostatin(SST)-expressing interneurons with Alzheimer's disease (AD) pathology (p=0.01) and cognitive decline score (p=0.03), and the microglial proportions are positively correlated with AD pathology (p=0.04) and APOÎµ E4 genotypes (p=0.01).
We estimated cell-type-specific multivariate models by regressing pseudo-bulk gene expression profiles on thousands of cis-regulatory variants. For the 45% of 26k genes, we identified 2-3 independent causal eQTLs per gene within 2Mb window (posterior probability > .9), and these independent variants correspond to 2-3 different cell types (correlation > .8). On a thousand individuals, we directly compared observed AD-related variables with gene expression variation imputed by cis-regulatory genotypes (TWAS). The significant examples include TOMM40 in pericytes associated with amyloid-beta, RAB2A in pericytes and IDI1 in excitatory neurons (L5/6) with tau protein level (FDR

PrgmNr 1290 - Accounting for coordinated regulatory activity allows to refine genetic regulation of gene expression in schizophrenia

View session detail

Author Block: **M. Alver**, N. Lykoskoufis, A. Ramisch, H. Ongen, E. T. Dermitzakis; Dept. of Genetic Med. and Dev., Univ. of Geneva, Geneva, Switzerland
Disclosure Block: **M. Alver:** None.
Schizophrenia (SCZ) is a heritable and highly polygenic neurodevelopmental disorder that oftentimes leads to a lifetime of chronic disability. While research has revealed numerous genomic loci predisposing to disease and identified dysregulation of gene expression in affected brain regions, causal mechanisms how genetic variants affect gene regulation in SCZ remains to be determined. Assessing the interplay between genetic variants, coordinated activity of regulatory elements (REs) and gene expression allows to fill that gap and gain insight into SCZ-specific gene regulatory mechanisms.
We used genotype, RNA-seq and ChIP-seq (histone mark H3K27ac) data collected within Human Brain Collection Core from post-mortem prefrontal cortex of 98 SCZ cases and 174 controls. To study the coordination of REs, we systematically measured interindividual correlation between proximal open chromatin peaks. Variability in nearby regulatory activity was structured into 10,938 and 10,376 well-delimited domains (cis-regulated domains, CRDs), in SCZ cases and controls, respectively, of which 42% were shared. Changes in 3D structure of the genome, rather than differential activity, were responsible for SCZ-specific CRDs. Differentially expressed genes were significantly enriched at differentially active CRDs (Fisher PTo interrogate deviations in regulatory signature patterns affecting gene expression in SCZ cases, we applied Bayesian Networks to infer the most likely causal relationship for QTLs affecting both CRDs and genes (QTL-CRD-gene triplets common to SCZ cases and controls (n=1,100)). While 2/3 of the triplets displayed the same direction of effect from QTL onto gene/CRD in both states, 1/3 of studied triplets showed change in directional effect from QTL variant onto gene expression (i.e., not mediated via associated CRD) in opposite state. These directional change-associated triplet genes clustered in molecular functions related to small GTPase binding, filopodium assembly and cellular lipid catabolic process, indicating changes in synaptic function and plasticity, and dendritic spine morphology in SCZ.
The established multi-level analysis accounting for coordinated activity between REs allowed to discriminate patterns of correlation structure between REs specific to SCZ, identify SCZ-specific QTLs as well as discover and refine perturbations in the molecular machinery underlying eQTL function in SCZ.

PrgmNr 1291 - Targeting transforming growth factor-Î² for treatment of osteogenesis imperfecta

View session detail

Author Block: **S. C. Sreenath Nagamani**[1,2], I-W. Song[1], D. Nguyen[1], I. Grafe[3], R. Sutton[4,2], F. Gannon[5], E. Munivez[1], M-M. Jiang[1], A. Tran[1], M. Wallace[6], P. Esposito[6], S. Musaad[7], E. Strudthoff[6], S. McGuire[6], M. Thornton[6], V. Shenava[8], S. Rosenfeld[8], R. Shypailo[9], E. Orwoll[10], B. Lee[4,2]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]Div. of Endocrinology, Diabetes, and Bone Diseases, Dept. of Med. and Ctr. for Hlth.y Aging, Univ. Clinic Dresden, Germany, and Ctr. for Regenerative Therapies Dresden, Univ. Dresden, Dresden, Germany, [4]Baylor Coll. Med., Houston, TX, [5]Dept.s of Pathology and Immunology and Orthopedic Surgery, Baylor Coll. Med., Houston, TX, [6]Orthopaedic Surgery, Univ. of Nebraska Med. Ctr., Childrenâ□□s Hosp. and Med. Ctr., Omaha, NE, [7]Dept. of Pediatrics-Nutrition, Baylor Coll. Med., Houston, TX, [8]Dept. of Orthopedic Surgery, Baylor Coll. Med., Houston, TX, [9]USDA/ARS Children's Nutrition Res. Ctr., Dept. of Pediatrics, Baylor Coll. Med., Houston, TX, [10]Dept. of Med., Bone and Mineral Unit, Oregon Hlth.& Sci. Univ, Portland, OR

Disclosure Block: **S.C. Sreenath Nagamani:** Grant/Contracted Research Support (External); Sanofi Genzyme.

Osteogenesis Imperfecta (OI), a Mendelian disorder of bone and connective tissue presents with low bone mass, recurrent fractures, bone deformities, as well as extraskeletal manifestations. To date, there are no disease-specific or FDA-approved therapies for OI. In preclinical studies, we have previously shown that excessive TGF-Î² signaling is a key driver of pathogenesis in OI. In this study, we evaluated TGF-Î² signaling in human OI bone and translated this discovery by conducting a phase 1 clinical trial of TGF-Î² inhibition in adults with OI. We first assessed signaling abnormalities in human OI bone using a multi-omic approach. Histology and RNASeq were performed on bones obtained from children with (n=10) and without (n=4) OI. Gene Ontology (GO) assay, gene set enrichment assay (GSEA), and Ingenuity Pathway Analysis (IPA) were used to identify key dysregulated pathways. Reverse-phase protein array, western blot, and immunohistochemistry were performed to evaluate changes at the protein level. We show that OI bone has woven structure, increased osteocyte density, high turnover, and reduced bone maturation. SMAD phosphorylation was the most significantly up-regulated GO molecular event. GSEA identified TGF-Î² to be the top activated pathway in OI. IPA showed that TGF-Î² was the most significant activated upstream regulator mediating the global changes identified in OI bone.

A phase 1 study with a single administration of two doses (1 mg/kg and 4 mg/kg) of a pan-anti-TGF-Î² neutralizing antibody, fresolimumab, was conducted in 8 adults with OI. The effect of fresolimumab on bone was assessed by measuring blood bone remodeling markers, and areal bone mineral density (aBMD). Treatment with fresolimumab was well-tolerated. Use of 4 mg/kg of fresolimumab was associated with suppressed bone remodeling compared to the 1 mg/kg dose. The percent change in osteocalcin (Ocn) was significantly different between the two doses (p=0.0045). Whereas there were no significant changes in C-terminus telopeptide (CTX) and Procollagen 1 Intact N-Terminal Propeptide (P1NP) between the two dose groups, similar to the results with Ocn, the percentage change of CTX from baseline was lower in the 4 mg/kg dose cohort. In 2 out of 4 individuals in each dose cohort, a robust increase in lumbar spine aBMD (6.7% and 8.6% in 1 mg/kg and 7.6% and 3% in 4 mg/kg dose) was observed compared to baseline. Our study, the first to comprehensively asses signaling abnormalities in human OI bone, confirms that TGF-Î² signaling is a driver pathogenic mechanism and that anti-TGF-Î² therapy could be a potential disease-specific therapy

PrgmNr 1383 - A cystic fibrosis lung disease modifier locus harbors tandem repeats associated with gene expression

View session detail

Author Block: **D. Roshandel**[1], S. Mastromatteo[1], C. Wang[1], J. Gong[1], B. Thiruvahindrapuram[1], W. W. L. Sung[1], Z. Wang[1], N. Panjwani[1], F. Lin[1], K. Keenan[1], L. Sun[2,3], J. M. Rommens[1], L. J. Strug[1,2,3]; [1]Genetics and Genome Biology Program, The Hosp. for Sick Children, Toronto, ON, Canada, [2]Dept. of Statistical Sci., Univ. of Toronto, Toronto, ON, Canada, [3]Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada

Disclosure Block: **D. Roshandel:** None.

Introduction: The largest genome-wide association study (GWAS) of cystic fibrosis (CF) lung disease identified two independent signals on chromosome 5 (Chr5:403,462-686,129) in 5â□□ and 3â□□ of a previously reported CF modifier gene, *SLC9A3*. The locus displays a high density of CpG islands and variable number of tandem repeats (VNTRs). The exact boundaries of these VNTRs have not been well defined as they often expand hundreds of bps difficult to capture by short-read sequencing.

Methods: We used long-read PacBio phased data and multiple alignment to identify the boundaries of common (> 2%) VNTRs in the region (N = 22). Association between the identified VNTRs and gene expression in the region (*AHRR*, *EXOC3*, *SLC9A3*, *CEP72* & *TPPP*) was then investigated using RNA-seq of CF nasal epithelia (NE; N = 46). Subsequently, the lengths of VNTRs were estimated in 10X Genomics (10XG) linked-read technology by dividing the number of reads aligned to the location of each VNTR by the average sequencing depth: we confirmed a high correlation between estimated lengths from short-read (10XG) sequencing and lengths from long-read (PacBio) sequencing in 53 subjects with both 10XG and PacBio. Therefore, the same strategy was used to estimate the VNTR lengths using short-read sequencing from the Genotype-Tissue Expression (GTEx) to investigate VNTR associations with gene expression in all 49 GTEx tissues. **Results:** At the locus, 54 VNTRs were identified. A VNTR in the last intron of *SLC9A3* overlapping a CpG island was associated with *SLC9A3* expression in NE (p = 5E-4) and forty GTEx tissues including lung (p = 8E-15). This VNTR was partially tagged by rs72711364, the top associated GWAS SNP 3â□□ of *SLC9A3* (Spearman correlation coefficient = 0.28). Its repeat sequence was â□□100bp including 7 CpGs. Subjects had 2-10 copies of the repeat (mean = 5) per haplotype. Another VNTR in the 3â□□ UTR of *TPPP* also overlapping a CpG island was associated with both *TPPP* (p = 9E-5) and *SLC9A3* (p = 2E-3) expression in NE, and multiple GTEx tissues. This VNTR had a 31 bp repeat sequence including 1-4 CpGs and multiple SNPs some changing CpG count. Subjects had 3-21 copies of the repeat per haplotype (mean = 9). The long/short version of this VNTR was perfectly tagged by C/T alleles of rs72703083, a GWAS SNP 5â□□ of *SLC9A3*. These two VNTRs together explained 22% and 9% of variation in *SLC9A3* expression in CF NE and GTEx lung, respectively. **Conclusion:** We used long-read sequencing to identify the precise boundaries of common VNTRs at the *SLC9A3* locus. Two of these VNTRs, tagged by genome-wide associated CF lung disease SNPs from two independent GWAS peaks in the region, account for almost a quarter of *SLC9A3* expression in the NE model of CF airway.

PrgmNr 1384 - Genome-wide somatic mutation rate maps uncover drivers of cancer

View session detail

Author Block: **M. A. Sherman**[1,2], A. Yaari[1], O. Priebe[1], P-R. Loh[3,2], B. Berger[1,2]; [1]Massachusetts Inst. of Technology, Cambridge, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Brigham and Women's Hosp. / Harvard Med. Sch., Boston, MA

Disclosure Block: **M.A. Sherman:** None.

Identifying cancer driver elements - genomic loci that can cause cell proliferation, tissue invasion, and immune evasion - is a critical challenge, particularly in the noncoding genome. We developed a probabilistic deep learning framework that integrates tissue-of-origin epigenetic patterns with cancer mutational spectra to learn a map of the neutral somatic mutation rate across the genome of a cancer of interest. These browsable maps allow dynamic retrieval of the mutation rate at any genomic locus and enable tests for positive selection by comparing the expected to observed mutation rate in a cancer cohort. We applied this approach to identify driver elements genome-wide in 2,583 tumor samples from the Pan-Cancer Analysis of Whole Genomes resource.

-

Predicted cryptic splice SNVs in the introns of tumor suppressor genes (TSGs) were observed significantly more often than expected under neutrality ($P=2.6Ã□10^{-5}$) with the majority (79.3%) outside the 20bp splice region bordering exons. These cryptic splice SNVs, which are usually overlooked by computational pipelines, accounted for 4.5% (95% CI: 1.6-7.3%) of driver SNVs in TSGs and recurrently disrupted 8 genes in 14 tumor types including *CIITA* (blood tumors) and *SMAD4* (pan-cancer). Of the 7 predicted cryptic splice SNV carriers with available tumor RNA-seq data, 6 had evidence of alternative splicing consistent with predictions.

-

A search for positive selection in noncoding regulatory elements revealed that the 5â□□ UTR of *TP53* carried a significant burden of indels (7 observed vs. 0.08 expected; $P=3.3Ã□10^{-12}$). These indels - all deletions - were larger than most indels (median: 17 bp) and recurrently disrupted regulatory sequences such as the transcription start site and a splice region within the UTR (*TP53* 5â□□ UTR spans multiple exons). The six *TP53* 5â□□ UTR mutation carriers with available tumor RNA-seq data exhibited 1-2 s.d. decreases in *TP53* expression ($P=1.6Ã□10^{-4}$, adjusted for tissue and copy number).

-

We also searched for driver mutations in infrequently mutated (â□□long-tailâ□□) genes in whole-exome and targeted sequencing of 14,018 tumors from 10 cancer types. Across these cancers, 1-5% of samples carried excess activating SNVs in oncogenes and 3-6% of samples carried excess protein-truncating SNVs in TSGs not associated with that cancer in recent pan-cancer driver gene databases. We identified 142 specific driver gene-tumor pairs not reported in these pan-cancer databases, detailing the rich, shared landscape of common and rare cancer driver genes. These results highlight the promise of genome-wide somatic mutation rate maps to uncover underappreciated coding and noncoding drivers of human cancers.

PrgmNr 1385 - Developmental Venous Anomaly: A genetic primer to *PIK3CA*-related neurological disease?

View session detail

Author Block: **D. A. Snellings**[1], R. Girard[2], R. Lightle[2], A. Srinath[2], S. Romanos[2], A. A. Ren[3], M. L. Kahn[3], I. A. Awad[2], D. A. Marchuk[1]; [1]Duke Univ., Durham, NC, [2]The Univ. of Chicago Med. and Biological Sci., Chicago, IL, [3]Univ. of Pennsylvania, Philadelphia, PA

Disclosure Block: **D.A. Snellings:** None.

Developmental Venous Anomalies (DVAs) are common vascular malformations present in up to 3% of the population and are generally considered to be benign. Despite being considered benign, DVAs have long been known to be directly associated with sporadic Cerebral Cavernous Malformations (CCMs)â□□a rarer, more aggressive type of vascular malformation that grow from the branches of a DVA. Recent work from us and others have shown that CCMs develop through co-occurring gain of function mutations in *PIK3CA* and *MAP3K3*. We hypothesize that DVA are the product of somatic *PIK3CA* mutation during developmental angiogenesis, creating field of mutant cells that may develop into a CCM upon a subsequent mutation in *MAP3K3*. To test this hypothesis, we used Droplet Digital PCR and targeted sequencing to detect somatic mutations in surgically resected sporadic CCMs in addition to paired samples of associated DVA discretely dissected from the DVA bed during surgery. We find that the *PIK3CA* mutation is present in both the DVA and the CCM, but that the *MAP3K3* mutation is exclusively present in the CCM suggesting that the *PIK3CA* mutation occurs first. Furthermore, we performed single-nucleus DNA sequencing on 3 CCM samples and found in each case that *PIK3CA* and *MAP3K3* mutations co-exist in the same clonal population of cells supporting a cell-autonomous synergy between *PIK3CA* and *MAP3K3* mutations. As *PIK3CA* is a commonly mutated oncogene and has been implicated in numerous cancer and non-cancer diseases, we wondered if DVA are also associated with diseases other than CCM. There are several clinical case reports that note a link between DVA and various brain tumors, however the strongest evidence comes from a recent study showing that DVA were present in 24.1% of a series of 162 diffuse intrinsic pontine gliomaâ□□a class of tumor reported to have *PIK3CA* mutations in up to 27% of cases. Notably, association with DVA is also present in the literature of other diseases related to PI3K activation including: hemimegalencephaly, focal cortical dysplasia, and Cowden syndrome. Taken together these data support a model where DVA serve as a genetic primer for the formation of CCM, glioma, and other *PIK3CA*-related disease.

PrgmNr 1386 - Investigation of a *HOXA* locus control region in early human craniofacial development

View session detail

Author Block: **A. Wilderman**[1], A. Hardy[2,3], T. Yankee[4], N. Glidden[1], E. Winchester[1], J. J. VanOudenhove[1,5], J. Cotney[1]; [1]UConn Hlth., Farmington, CT, [2]UConn, Farmington, CT, [3]Univ. de Paris, CNRS, Inst. Jacques Monod, Paris, France, [4]UConn Hlth., farmington, CT, [5]Yale Univ. Sch. of Med., Dept. of Hematology, New Haven, CT

Disclosure Block: **A. Wilderman:** None.

Defects in embryonic patterning resulting in craniofacial abnormalities account for approximately 1/3 of birth defects. The regulatory programs that build and shape the face require precisely controlled spatiotemporal gene expression, achieved through tissue-specific enhancers. Large regions with coactivation of enhancer elements and co-regulation of multiple genes, referred to as superenhancers, are important in determining cell identity and perturbation could result in developmental defects. Building upon a previously published epigenomic atlas of human embryonic craniofacial tissue in which we identified over 75,000 putative embryonic craniofacial enhancer regions, we have identified 581 superenhancer regions, including 36 which fall in completely non-coding regions and not directly adjacent to known craniofacial relevant genes. To demonstrate the utility of this data for the understanding of craniofacial development and the etiology of craniofacial abnormalities, we focused on a craniofacial-specific superenhancer in a ~600kb non-coding region located between *NPVF* and *NFE2L3*. This region harbors over 100 individual putative craniofacial enhancer segments and 7 in vivo validated craniofacial enhancers from primary craniofacial tissue as well as strong enhancer activation signatures in a culture model of cranial neural crest cell (CNCC) development. However, none of the directly adjacent genes have been implicated in neural crest specification, craniofacial development, or abnormalities. To identify potential regulatory targets of this super enhancer region, we characterized three-dimensional chromatin structure of this region in CNCCs and mouse embryonic craniofacial tissues using multiple techniques (4C-Seq, HiC). We identified long range interactions that exclude most intervening genes and specifically target the anterior portion of the *HOXA* gene cluster located 1.2 to 1.8 Mb away. We demonstrate the specificity of the enhancer region for regulation of anterior *HOXA* genes through CRISPR/Cas9 editing of human embryonic stem cells. Mice homozygous for deletion of the superenhancer confirm the specificity of the enhancer region and demonstrate that the region is essential for viability. At fetal stages homozygotes develop at the same rate as heterozygous and wild type littermates but die at P0-P1 and have high penetrance of orofacial clefts that phenocopy previously described *Hoxa2* knockout mice. This evidence suggests we have identified a critical non-coding locus control region that specifically regulates anterior *HOXA* genes and whose deletion could be pathogenic in human patients.

Platform Sessions

PrgmNr 1004 - Revisiting the out of Africa event with a deep learning approach

View session detail

Author Block: **M. Mondal**[1], F. Montinaro[2], V. Pankratov[1], B. Yelmen[1], L. Pagani[3]; [1]Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [2]Dept. of Biology-Genetics, Univ. of Bari, Bari, Italy, [3]Univ. of Padova, Padova, Italy
Disclosure Block: **M. Mondal:** None.
Anatomically modern humans evolved around 300 thousand years ago in Africa. Modern humans started to appear in the fossil record outside of Africa about 100 thousand years ago though other hominins existed throughout Eurasia much earlier. Recently, several researchers argued in favourof a single out of Africa event for modern humans based on whole-genome sequences analyses. However, the single out of Africa model is in contrast with some of the findings from fossil records, which supports two out of Africa, and uniparental data, which proposes a back to Africa movement. Here, we used a deep learning approach coupled with Approximate Bayesian Computation and Sequential Monte Carlo to revisit these hypotheses from the whole genome sequence perspective. Our results support the back to Africa model over other alternatives. We estimated that there are two successive splits between Africa and out of African populations happening around 60-90 thousand years ago and separated by 13-15 thousand years. One of the populations resulting from the more recent split has to a large extent replaced the older West African population while the other one has founded the out of Africa populations.

PrgmNr 1005 - An analysis of population copy number variation in sub-Saharan African genomes

View session detail

Author Block: **E. Wiener**[1,2], L. Cottino[1,2], D. Jakubosky[3], A. Macleod[4], H. Noyes[5], O. Nyangiri[6,7], A. Krause[1], S. Hazelhurst[8,2], Z. Lombard[1], H3Africa Consortium; [1]Div. of Human Genetics, Natl. Hlth.Lab. Service & Sch. of Pathology, Faculty of Hlth.Sci., Univ. of Witwatersrand, Johannesburg, South Africa, [2]Sydney Brenner Inst. for Molecular BioSci., Faculty of Hlth.Sci., Univ. of the Witwatersrand, Johannesburg, South Africa, [3]Dept. of BioMed. Informatics, Univ. of California San Diego, La Jolla, CA, [4]Wellcome Trust Ctr. for Molecular Parasitology, Glasgow, United Kingdom, [5]Ctr. for Genomic Res., Univ. of Liverpool, Liverpool, United Kingdom, [6]Coll. of Vet. Med., Animal Resources and Biosecurity, Makerere Univ., Kampala, Uganda, [7]Epidemiology and Demography Dept., KEMRI-Wellcome Trust Res. Programme, Kilifi, Kenya, [8]Sch. of Electrical and Information Engineering, Univ. of Witwatersrand, Johannesburg, South Africa

Disclosure Block: **E. Wiener:** None.

Introduction Copy number variation (CNV) is responsible for a large component of normal human variation and has been implicated in the cause/genetic aetiology of several rare diseases. Population reference databases containing CNV information from all global populations is critical in disease genetics research, but current resources lack diversity, especially from the African continent. This makes such databases of limited use in studies looking at genetic diseases in African individuals. This study therefore aims to address this knowledge gap by producing a map of CNV using whole-genome data from several, previously unstudied African populations

Methods 1027 high coverage whole genome sequences obtained from individuals across west, central, southern and east Africa, were analysed using Manta and Graphtyper2. Additionally, 919 of the samples were also analysed using Genome STRiP to detect multi-allelic CNV. Quality control specific to each tool was performed in order to achieve high quality variant call sets.

Results 56 816 CNVs were detected by the Manta pipeline, consisting of 44 671 deletions and 12 145 duplications. Due the ability of Manta to detect small variants (5% allele frequency. 50% of these variants are novel compared to 27% of the remaining variants >100bp. Overall, 32% of the variants identified were novel. A comparison between central, west, east and southern African regions yielded a number of variants unique to each region. We find deletions tend to have lower allele frequencies compared to duplications. The majority of variants were found in the non-coding genome, with only 8% of variants overlapping coding transcripts. An additional 5% of variants overlapped regulatory features. Genome STRiP detected 3991 multi-allelic variants with 99% having a copy number between 3-20. There were also variants with copy numbers greater than 20, some of which appear to be incidences of excessive runaway duplications not previously described.

Conclusion The amount of novel variation found demonstrates the importance of including African individuals from multiple African regions when producing reference databases and the rich genomic diversity of African genomes. Work is currently being performed to combine the full Genome STRiP and Manta call sets to produce a robust combined dataset. The variant database produced in this study will provide a valuable resource as a reference of normal CNV for the study of diseases in African populations.

PrgmNr 1006 - Integrative genomic analyses identify key interethnic differences in immune response to malaria

View session detail

Author Block: **Y. Idaghdour**[1], W. Abdrabou[1], M. Dieng[1], A. Diawara[1], S. Samuel[2], D. Almojil[1], S. SombiÃ©[2], N. Henry[2], D. Kargougou[2], V. Manikandan[1], I. Soulama[2]; [1]New York Univ. Abu Dhabi, Abu Dhabi, United Arab Emirates, [2]Ctr. Natl. de Recherche et de Formation sur le Paludisme, Ouagadougou, Burkina Faso

Disclosure Block: **Y. Idaghdour:** None.

Host responses to infection with the malaria parasite *P. falciparum* vary between individuals for reasons that are poorly understood. Here we reveal metabolic perturbations as a consequence of malaria infection in children and identify an immunosuppressive role of endogenous steroid production in the context of *P. falciparum* infection. We perform metabolomics on matched samples from children from two ethnic groups in West Africa, before and after infection with seasonal malaria. Analyzing 306 global metabolomes we identify 92 parasitemia-associated metabolites with impact on the host adaptive immune response. Integrative metabolomic-transcriptomic and causal mediation-moderation analyses reveal an infection-driven immunosuppressive role of parasitemia-associated pregnenolone steroids on lymphocyte function and the expression of key immunoregulatory lymphocyte genes in the Gouin ethnic group. In children from the less malaria-susceptible Fulani ethnic group we observe opposing responses upon infection, consistent with the immunosuppressive role of endogenous steroids in malaria. These findings advance our understanding of *P. falciparum* pathogenesis in humans and identify potential new targets for antimalarial therapeutic interventions.

PrgmNr 1007 - GWAS of complex traits in a multi-population African cohort

View session detail

Author Block: **M. Hansen**[1], S. A. Tishkoff[2], J. Lachance[3], A. Ranciaro[1], M. Kim[4], S. Raj[5], S. Fan[6], W. R. Beggs[1]; [1]Univ Pennsylvania, Philadelphia, PA, [2]Univ. of Pennsylvania Sch. of Med., Philadelphia, PA, [3]Georgia Inst. of Technology, Atlanta, GA, [4]Fred Hutchinson Cancer Res. Ctr., Seattle, [5]Cornell Univ, Lexington, MA, [6]Perelman Sch. of Med., Philadelphia, PA

Disclosure Block: **M. Hansen:** None.

The diversity among present-day African populations is the result of a deep and complex history of admixture, migrations, and regional adaptations to local environments and diseases. Little is known about the impact of this evolutionary history on the genetics underlying complex traits. Here I present recent work on genetic associations for a panel of anthropometric, cardiovascular, and metabolic biomarker measurements paired with dense genotyping data. For some traits, the variation among populations is expected to reflect local adaptations, such as short stature in western Cogo rainforest hunter-gatherers. The study cohort of several thousand individuals is drawn from an ancestrally diverse set of populations from western, eastern, and southern sub-Saharan Africa. Populations include current or recent hunter-gatherers, traditional agriculturalists, and semi-nomadic pastoralists, from rural regions of Cameroon, Nigeria, Ethiopia, Kenya, Tanzania, and Botswana. For many of these traits, this marks the first genotype/phenotype analysis to include these ethnic groups. The high degree of population structure presents both challenges and opportunities for genetic analysis. Genetic structure analysis indicates genetic clustering by geographic location, language family, and regional hunter-gatherer lineages. Examples include the hunter-gathers from the Serengeti, western Congo, and Kalahari, and clusters that correlate with Niger-Congo, Afroasiatic, and Nilo-Saharan language families. We observe substantial population-level variation for many traits, such as height, skin pigmentation, and blood pressure. The proportion of the trait variance that is due to the genetic population structure varies by trait and tends to be greater for anthropometric traits like height and skin pigmentation than for metabolic biomarkers like LDL. From genotype/phenotype association tests we find numerous independent associations at genome-wide significance for several traits, including circulating triglyceride levels and BMI. The population structure of the total additive genetic effects is also examined. European GWAS associations replicate poorly in this African cohort, while associations discovered in the African cohort show comparatively better replication in Europeans.

PrgmNr 1008 - Genotype-by-infection interactions: Single cell RNASeq profiling of in-vivo host immune response to malaria reveals cell type and infection-specific eQTLs

View session detail

Author Block: **O. Bayaraa**[1], M. Dieng[1], J. Ryou[1], V. Manikandan[1], W. Abdrabou[1], S. SermÃ©[2], S. SombiÃ©[2], A. Barry[2], S. Coulibaly[2], A. Diarra[2], M. Arnoux[1], A. B. Tiono[2], S. Sirima[2], A. Diawara[1], I. Soulama[2], Y. Idagdhour[1]; [1]New York Univ. Abu Dhabi, Abu Dhabi, United Arab Emirates, [2]Ctr. Natl. de Recherche et de Formation sur le Paludisme, Ouagadougou, Burkina Faso
Disclosure Block: **O. Bayaraa:** None.
The disease burden of malaria remains a significant global public health challenge. *Plasmodium falciparum* is responsible for more than 99% of malaria cases in Africa and for >400,000/year malaria-related deaths worldwide. Inter-individual differences in susceptibility to malaria is multifactorial and has a significant heritable component but our understanding of the effect of infection on gene regulation of immune response at the transcriptional remains very limited. Here we use longitudinal matched sampling, single cell RNAseq profiling of PBMCs and whole-genome sequencing data of malarial children before and after natural *P. falciparum* infection in Banfora, Burkina Faso, West Africa. In total, we generated ~90,000 single cell RNASeq profiles and identified PBMC cell types affected by infection. Single cell RNASeq eQTL analysis revealed cell type specific eQTLs and genome-wide significant genotype-by-infection interaction effects implicating key immune genes. These results provide the first genome-wide picture of host in vivo regulatory variation events in malaria at the single cell level and highlight the implication of regulatory interaction effects in modulating host immune response in-vivo.

PrgmNr 1009 - Returning secondary genetic findings: Provider perspective in Africa

View session detail

Author Block: **A. Oladayo**[1], S. Prochaska[1], T. Busch[1], W. Adeyemo[2], L. Gowans[3], E. Mekonen[4], W. Awotoye[1], A. Alade[5], A. Butali[1], A. Prince[1], J. C. Murray[6], A. A. Adeyemo[7], P. Mossey[8]; [1]Univ. of Iowa, Iowa City, IA, [2]Univ. of Lagos/ Lagos Univ. Teaching Hosp., Akoka, Nigeria, [3]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [4]Addis Ababa Univ., Addis Ababa, Ethiopia, [5]Univ. of Iowa, Iowa, IA, [6]U of Iowa, Iowa City, IA, [7]NIH, Bethesda, MD, [8]Univ. of Dundee, Dundee, United Kingdom
Disclosure Block: **A. Oladayo:** None.
Objective: Previous research has shown that lack of resources and knowledge significantly impact the return of genomic test results. However, not much is known about the level of expertise and knowledge of clinicians providing cleft care in Africa on genetic diseases, despite the vast genetic diversity in this population.
Methods: Providers in participating cleft-craniofacial clinics in Ethiopia, Ghana, and Nigeria were sent the link to a 63-question online survey. This survey assessed the providers' experience with genetic testing, genetics education and return of genetic results, provider knowledge, clinician comfort with returning results, available resources to assist with genomic findings, and potential barriers.
Results: As of June 2nd, 2021, 246 providers completed the survey. Only 2% had been involved in the delivery of Exome or Genome sequencing; 78.6% had no formal genetic education, 49.6% agreed that all secondary findings should be disclosed to patients. Regarding the comfort level, 89.4% were somewhat to extremely comfortable discussing genetic risk factors with patients, and 81.8% were somewhat to extremely comfortable with returning genetic results. Sixty-three percent believed that resources were currently available to enable them to access needed genetic information.
Conclusion: Providers were aware that genetic testing could help in the clinical management of diseases from the returned responses. However, the lack of knowledge about genomic medicine, uncertain clinical utility, and lack of available resources were cited as barriers that significantly impacted incorporating genetic testing into their practice. Data collection is ongoing and will continue till July 31st, 2021. This is the first Ethical, Legal, and Social Implications (ELSI) study to document the knowledge and comfort level of cleft providers in Africa. This study will help determine the most beneficial information to equip providers with the return of secondary genetic findings.

PrgmNr 1012 - Multi-trait Bayesian genome-wide association study identifies novel endometrial cancer risk loci at 7q22.1, 8q24.3 and 16q22.2

View session detail

Author Block: **T. A. O'Mara**[1], X. Wang[1], P. Kho[1], C. Bafligil[2], D. G. Evans[2,3], E. J. Crosbie[2,3], A. B. Spurdle[1], D. M. Glubb[1], Endometrial Cancer Association Consortium; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]Univ. of Manchester, Manchester, United Kingdom, [3]Manchester Univ. NHS Fndn. Trust, Manchester, United Kingdom

Disclosure Block: **T.A. O'Mara:** None.

Endometrial cancer is the most common gynecological tumor in developed countries, and its prevalence and mortality are increasing worldwide. Current genome-wide association studies (GWAS) of endometrial cancer, performed by the Endometrial Cancer Association Consortium, have identified 16 genetic risk regions for this disease. To identify additional risk variants for endometrial cancer, we leveraged summary statistics from the largest GWAS meta-analysis for endometrial cancer performed to date (12,906 cases and 108,979 controls) and publicly available GWAS datasets of endometrial cancer risk factors using a recently developed multi-trait Bayesian GWAS (bGWAS) framework. Using LD score regression, we assessed the genetic correlation between >2,500 phenotypes and endometrial cancer risk to identify potential endometrial cancer risk factors. Significantly genetically correlated phenotypes (FDRThe bGWAS findings suggest that the three novel loci affect endometrial cancer risk through altered sex-hormone levels or effects on obesity. Consistent with this hypothesis, several genes with established roles in these pathways (*CYP11B1*, *CYP3A7*, *IRX3* and *IRX5*) were prioritized as candidate endometrial cancer risk genes by interrogation of quantitative trait loci data and chromatin looping assays in endometrial cell lines. The findings of this study identify additional opportunities for hormone therapies and further support weight loss to reduce the risk of developing endometrial cancer.

PrgmNr 1013 - Novel basal cell carcinoma susceptibility loci identified from a multiethnic genome-wide meta-analysis

View session detail

Author Block: **H. Choquet**[1], J. Yin[1], Y. Kim[2], T. J. Hoffmann[3], E. Jorgenson[4], M. M. Asgari[5]; [1]Kaiser Permanente, Oakland, CA, [2]Harvard Med. Sch. and Harvard Pilgrim Hlth.Care Inst., Boston, MA, [3]Inst. for Human Genetics, Univ. of California, San Francisco (UCSF), San Francisco, CA, [4]Regeneron Genetics Ctr., Tarrytown, NY, [5]Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA
Disclosure Block: **H. Choquet:** None.
Purpose: Basal cell carcinoma (BCC) is one of the most common malignancies worldwide and has a moderate genetic component with an array-heritability estimate of up to 17.0%. Previously published genome-wide association studies (GWAS) have reported more than 70 loci associated with BCC, explaining up to 11.0% of BCC heritability, suggesting that additional loci remain to be discovered. Most published studies have exclusively utilized European descent populations, and the impact of genetic factors on BCC risk in a multiethnic population has not been explored. **Methods:** We conducted GWAS analyses, followed by a meta-analysis, including 526,432 individuals (34,616 BCC cases and 491,816 controls) from two cohorts: the Kaiser Permanente Genetic Epidemiology Research in Adult Health and Aging (GERA) and the UK Biobank (UKB). First, a logistic model GWAS for BCC was carried out separately in each ethnic group of GERA (non-Hispanic whites, Hispanic/Latinos, East Asians, and African-Americans), and UKB (European, South Asian, African British, and mixed Ancestries), adjusting for age, sex, and ancestry principal components. Then, the results from each ethnic group, and each cohort were meta-analyzed. We tested novel loci discovered in the combined multiethnic meta-analysis in an external replication cohort. To prioritize genes and biological pathways, and highlight gene-set and tissue/cell enrichments within the associated loci identified, we used DEPICT (Data-driven Expression-Prioritized Integration for Complex Traits) integrative tool. **Results:** Our multiethnic GWAS meta-analysis identified 78 genome-wide significant (*P*â□□8) BCC-associated loci, of which 20 were novel. Most of them replicated in an external independent sample at Bonferroni significance with the same direction of effect. Identified novel loci are implicated in the normal cell function (*SENP7*), adhesion of platelets and other cell types to the extracellular matrix (*ITGA2*), or DNA repair (*ATM*). Interrogation of loci uncovered by our multiethnic meta-analysis GWAS using DEPICT integrative tool also prioritized novel biological pathways underlying BCC. **Conclusions:** In this work, we present the largest and most ethnically diverse genetic study of BCC susceptibility conducted to date to our knowledge. Study findings provide new insight into the genetic basis of BCC susceptibility and may help identify individuals at higher BCC risk.

PrgmNr 1014 - A fine-mapping study of over 254,000 Asian and European descendants identifies putative susceptibility genes for colorectal cancer

View session detail

Author Block: **Z. Chen**[1], X. Guo[1], J. Huyghe[2], P. Law[3], C. Fernandez-Rozadilla[4], M. Timofeeva[5], M. Thomas[2], S. Schmit[6], V. DÃez-Obrero[7], L. Hsu[2], J. Long[1], V. Moreno[7], G. Casey[8], M. Dunlop[9], S. B. Gruber[10], I. Tomlinson[9], U. Peters[2], R. Houlston[3], W. Zheng[1]; [1]Div. of Epidemiology, Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [3]The Inst. of Cancer Res., London, United Kingdom, [4]Inst. de Investigacion Sanitaria de Santiago - IDIS, Santiago de Compostela, Spain, [5]Univ. of Southern Denmark, Odense, Denmark, [6]Cleveland Clinic, Cleveland, OH, [7]Bellvitge BioMed. Res. Inst., Barcelona, Spain, [8]Dept. of Publ. Hlth.Sci., Univ. of Virginia, Charlottesville, VA, [9]Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom, [10]USC, Los Angeles, CA
Disclosure Block: **Z. Chen:** None.

Over 150 genetic susceptibility loci have been identified for colorectal cancer (CRC) through genome-wide association studies (GWAS). To uncover potential independent association signals and identify putative target genes in these loci, we conducted a fine-mapping analysis using GWAS summary statistics data from 72,272 subjects of East Asian (EAS)- and 182,518 subjects of European-ancestry (EA). We performed stepwise conditional analyses by race using the GCTA-COJO with adjustment for the top association signal in each region. Results from EAS and EA were combined using a fixed-effects model. To identify race-specific associations, similar analyses were performed with adjustment for the race-specific top association signal in each region. At conditioned *P* -6, we identified 219 independent risk signals, including 142 from combined analyses of EA and EAS, 9 from EAS analyses, and 68 from EA analyses. Using a conditioned *P* value within two orders of magnitude of the most statistically significant variant for each signal, we identified 6,131 credible causal variants (CCVs). Using a modified INQUSIT pipeline to score each gene-CCV pair by integrating functional genomic data, we identified 168 potential target genes. We conducted cis-eQTL analysis using transcriptome data derived from colon transverse tissues of EA from GTEx and of tumor-adjacent normal tissue of EAS. At Bonferroni-corrected *P* P P

PrgmNr 1015 - Assessment of genetic susceptibility to multiple primary tumors through whole-exome sequencing in two large cohorts

View session detail

Author Block: **T. B. Cavazos**[1], L. Kachuri[1], R. E. Graff[1], J. L. Nierenberg[1], K. K. Thai[2], S. Alexeeff[2], S. Van Den Eeden[2], Regeneron Genetics Center, T. Hoffmann[1], E. Ziv[1], L. Habel[2], E. Jorgenson[3], L. C. Sakoda[2], J. S. Witte[1]; [1]Univ. of California, San Francisco, San Francisco, CA, [2]Kaiser Permanente, Oakland, CA, [3]Regeneron, Tarrytown, NY
Disclosure Block: **T.B. Cavazos:** None.
Approximately 10-20% of individuals diagnosed with one cancer will face another cancer diagnosis in their lifetime. Multiple cancer diagnoses will likely become more common with earlier detection and improving survival. While some pleiotropic drivers of hereditary cancer syndromes, such as BRCA1/2 and TP53, have been discovered, the genetic factors contributing to the development of multiple primary tumors within single individuals are not well understood. To characterize the genetic susceptibility to multiple malignancies we conducted a pan-cancer, whole-exome sequencing study of two large prospective cohorts: Kaiser Permanente (KP) and the UK Biobank (UKB). The former cohort included 3,111 individuals with two or more primary cancers and 3,126 cancer-free controls, and the latter included 2,126 cases and 173,779 controls. In KP, we used SAIGE to conduct gene-based tests of predicted functional/pathogenic rare (MAF significant associations after Bonferroni correction (p

PrgmNr 1016 - Validation of the BOADICEA multifactorial breast cancer risk prediction model in a large prospective cohort study

View session detail

Author Block: **X. Yang**[1], M. Eriksson[2], K. Czene[2], A. Lee[1], G. Leslie[1], J. Wang[1], J. Dennis[1], L. Dorling[1], S. Carvalho[1], N. Mavaddat[1], J. R. Simard[3], M. K. Schmidt[4,5], D. F. Easton[1,6], P. Hall[2,7], A. C. Antoniou[1]; [1]Ctr. for Cancer Genetic Epidemiology, Dept. of Publ. Hlth.and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, [2]Dept. of Med. Epidemiology and Biostatistics, Karolinska Inst.t, Stockholm, Sweden, [3]Univ. Laval, Quebec, QC, Canada, [4]Dept. of Clinical Genetics, Leiden Univ. Med. Ctr., Leiden, Netherlands, [5]Div. of Molecular Pathology, The Netherlands Cancer Inst. - Antoni van Leeuwenhoek Hosp., Amsterdam, Netherlands, [6]Ctr. for Cancer Genetic Epidemiology, Dept. of Oncology, Univ. of Cambridge, Cambridge, United Kingdom, [7]Dept. of Oncology, South Gen. Hosp., Stockholm, Sweden

Disclosure Block: **X. Yang:** None.

Background The BOADICEA breast cancer (BC) risk prediction model has been extended to incorporate questionnaire-based risk factors (QRFs), mammographic density (MD), explicit family history (FH) and the effects of known rare and common BC genetic susceptibility variants. Here, for the first time, we assessed its clinical validity in predicting the 5-year BC risks, using an independent, prospective cohort. **Methods** We used data from the Swedish KARMA screening cohort, which included 66,415 women (816 incident BCs) without a previous cancer diagnosis, recruited between 2011 and 2013. All participants had information on QRF, FH, and categorical MD (BIRADS). A sub-cohort of 15,502 women (676 incident BCs) had a validated BC polygenic risk score (PRS) based on 313-SNPs, and 5,693 women (280 incident BCs) had gene-panel testing information for the 5 major BC susceptibility genes: *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, and *ATM*. We assessed the model calibration and discriminatory ability in predicting 5-year BC risks, using a weighted cohort analysis approach. Calibration was assessed using the ratio of expected versus observed BC cases in deciles of predicted risk and the calibration slope. The discriminatory ability was assessed using the area under the curve (AUC). **Results** BOADICEA was well calibrated in the full cohort on the basis of QRF, MD and FH (calibration slope=0.98, 95%CI: 0.96-1.00). In the subcohort with PRS information, the largest discrimination was achieved by the PRS (AUC=0.665, 95%CI:0.644-0.686), followed by QRF (AUC=0.626, 95%CI:0.605-0.647), FH (AUC=0.612, 95%CI:0.591-0.633) and MD (AUC=0.626, 95%CI: 0.605-0.646). When all four risk factors were considered, AUC was 0.685 (95%CI:0.666-0.705). The model was well calibrated in deciles of predicted risks (overall calibration slope 0.98, 95%CI:0.96-1.00). AUC was 0.688 (95%CI:0.651-0.724) in pre-menopausal women and 0.665 (95%CI:0.641-0.690) in post-menopausal women. In the subcohort with gene-panel testing information, addition of pathogenic variant status in the 5 BC susceptibility genes improved discrimination with an AUC=0.696 (95%CI: 0.660-0.732) and the model remained well calibrated in deciles of predicted risk (overall calibration slope 0.97, 95%CI: 0.95-0.99). **Conclusion** The study shows that the multifactorial BOADICEA model (version 5.0) provides valid BC risk predictions, and that incorporating all risk factors provides significantly better discrimination than any one factor. BOADICEA, which is implemented in the CanRisk tool (www.canrisk.org) for use by healthcare professionals, provides a powerful basis for personalised decisions on disease prevention and screening.

PrgmNr 1017 - High resolution functional mapping credible causal variants to target genes at 54 common melanoma susceptibility loci

View session detail

Author Block: **R. Thakur**[1], M. Xu[1], L. Jessop[1], H. Sowards[1], T. Zhang[1], T. Myers[1], M. E. Johnson[2], A. D. Wells[2], A. Chesi[3], S. F. Grant[4], M. I. Iles[5], M. T. Landi[1], M. Law[6], J. Choi[1], Melanoma Meta-Analysis Consortium, M. J. Machiela[1], K. M. Brown[1]; [1]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD, [2]Ctr. for Spatial and Functional Genomics, Div.s of Human Genetics and Endocrinology, Childrenâ□□s Hosp. of Philadelphia Res. Inst. and Dept. of Pediatrics, Perelman Sch. of Med., Univ. of Pennsylvania Philadelphia, Philadelphia, PA, [3]Ctr. for Spatial and Functional Genomics, Div.s of Human Genetics and Endocrinology, Childrenâ□□s Hosp. of Philadelphia Res. Inst. and Dept. of Pediatrics, Perelman Sch. of Med., Univ. of Pennsylvania Philadelphia, Wallingford, PA, [4]Ctr. for Spatial and Functional Genomics, Div.s of Human Genetics and Endocrinology, Childrenâ□□s Hosp. of Philadelphia Res. Inst. and Dept. of Pediatrics, Perelman Sch. of Med., Univ. of Pennsylvania Philadelphia,, Philadelphia, PA, [5]Leeds Inst. for Data Analytics, Univ. of Leeds, Leeds, United Kingdom, [6]Statistical Genetics, QIMR Berghofer Med. Res. Inst., Brisbane, Australia

Disclosure Block: **R. Thakur:** None.

Quantitative Trait Locus (QTL) studies have been useful for nominating candidate causal genes (CCGs) for genome-wide association study (GWAS) loci. However, melanocyte-specific expression (eQTL) and methylation (mQTL) QTLs nominate CCGs for only 37% of all known melanoma GWAS loci by colocalization. To complement these QTL data and conduct variant to gene mapping at a higher resolution, we performed region-focused chromatin conformation capture assay (capture-C) to identify physical interactions between Credible Causal Variants (CCV) and target genes in primary human melanocytes. We baited the regions of association for all independent signals from a recent melanoma GWAS (n=36,760 cases; Landi et al. 2020) and capture-C interactions were called using HiCUP and CHiCAGO. Capture-C data identified physical interactions between CCVs and CCGs at 57% of the loci with no QTL support; collectively QTL and capture-C data nominated CCGs for 94% of loci. We performed sequencing assay for transposase-accessible chromatin (ATAC-seq) in primary melanocytes and observed interactions between CCVs located within open chromatin to CCG at 70% of the loci. These findings were in concordance with the ATAC-seq data from melanoma cell lines and interactions from CCVs in melanoma open chromatin regions were observed for 60% of loci. Furthermore, we annotated CCV to CCG interactions using the available ChromHMM data from primary melanocytes and melanoma cell lines and observed that these interactions formed enhancer-promoter loops at 50% and 38% of loci in melanocytes and melanoma cell lines respectively. Biological pathway enrichment analyses of capture-C-nominated CCGs identified embryonic development, aryl hydrocarbon receptor signaling, interferon signaling, and DNA repair pathways, including multiple targets involved in the neural crest lineage. Notably, we observed interactions from two independent loci (separated by ~1.6Mb) to the promoter of *SOX4*, a gene that regulates neuronal differentiation. The first independent locus near *HDGFL1* showed multiple loops from the region of association to the *SOX4* promoter, one of which included CCV rs16886790. The second locus, located in a melanocyte enhancer within the *CDKAL1* gene, also showed direct interaction between a region harboring multiple CCVs and the *SOX4* promoter. Integration of these results with MPRA and CRISPR screening data are currently underway. In summary, we developed a high-throughput cell-type specific analysis framework for high-resolution functional mapping CCVs to CCGs and highlighted the added utility of capture-C for identifying high-confidence CCGs at melanoma susceptibility loci.

PrgmNr 1020 - Analysis of PI3K-AKT-MTOR spectrum disorders through deep genomic and functional models reveals new clinical insights and distinct molecular pathomechanisms

View session detail

Author Block: **F. Pirozzi**[1], M. Berkseth[1], R. Shear[1], G. Ruggeri[1], N. Horsley[1], L. Gonzalez[2], A. E. Timms[1], J. Sulc[1], E. Pao[1], N. Oyama[1], F. Forzano[3], V. Conti[4], R. Guerrini[4], E. S. Doherty[5], W. B. Dobyns[6], S. C. Saitta[7], E. Novotny[8], G. M. Mirzaa[1]; [1]Seattle Children's Res. Inst., Seattle, WA, [2]John Hopkins Univ., Baltimore, MD, [3]Guy's and St Thomas NHS Fndn. Trust, London, United Kingdom, [4]A. Meyer Children's Hosp., Florence, Italy, [5]Carilion Clinic Children's Hosp., Roanoke, VA, [6]Univ. of Minnesota, Minneapolis, MN, [7]David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [8]Seattle Childrens Hosp., Seattle
Disclosure Block: **F. Pirozzi:** None.
Megalencephaly (MEG), Hemimegalencephaly (HMEG) and Focal Cortical Dysplasia (FCD) are neurodevelopmental disorders characterized by brain overgrowth and cortical abnormalities, associated with significant pediatric morbidity and mortality including epilepsy, autism and intellectual disability. Gain and loss of function mutations in the PI3K-AKT-MTOR pathway have been identified in this spectrum, with variable levels of mosaicism and distribution of causal mutations. We aimed to better define the genotype-phenotype correlation among the most common hotspot mutations in the pathway. We analyzed a cohort of 58 patients (144 samples, 114 epilepsy brain tissue) and tested them using droplet digital PCR (ddPCR). We were able to solve 29% of our cohort, with a diagnostic yield of 90% in HMEG and 24% in FCD. Interestingly, individuals with FCD mostly had mutations in *MTOR*, while those with HMEG had *PIK3CA* mutations. Our data highlighted an inverse correlation between age of onset for seizure and mutational burden (level of mosaicism), with individuals presenting a Variant Allele Fraction (VAF) >5% having seizure at day of life 1-30, and individuals with VAFin vitro using patient-derived and genome edited induced pluripotent stem cells carrying the common $PIK3CA^{H1047R}$ and $MTOR^{T1977I}$ mutations and differentiated them to Neuronal Progenitor Cells (NPCs) and cerebral organoids. We performed functional assays including population doubling time, analysis of cellular morphology and size, differentiation and cortical layering.Our results show overlapping and exclusive cellular phenotypes in *PIK3CA* and *MTOR* mutant cell lines. Specifically, both mutant lines had increased NPCs proliferation and hypertrophy, albeit *PIK3CA* mutants displayed the most severe phenotype. *MTOR* and *PIK3CA* mutant organoids had an average area at least double that of controls, recapitulating MEG *in vitro*.Eventually, we treated NPCs with 3 different MTOR pathway inhibitors to investigate their potential in rescuing the cellular phenotypes identified. Inhibition of MTOR activity *in vitro* effectively reduced hypertrophy and hyperplasia, providing a proof of concept for novel therapies for individuals with FCD and HMEG.In summary, our results show that distinct mechanisms underlie *PIK3CA*- and *MTOR*- related MEG and FCD during human embryonic development, and that inhibition of MTOR pathway might be sufficient to restore cellular homeostasis. This work will help characterize mechanisms of MEG and FCD caused by mutations in MTOR pathway and guide the design of future clinical trials.

PrgmNr 1021 - Skewed X-chromosome inactivation in unsolved neurodevelopmental disease cases can guide re-evaluation for X-linked genes

View session detail

Author Block: **C. Giovenino**[1], L. Pavinato[1,2], S. Trajkova[1], S. Cardaropoli[3], V. Pullano[1], S. Carestiato[1], P. Salmin[4], D. Carli[3], A. Mussa[3], S. De Rubeis[5,6,7,8], J. D. Buxbaum[9,5,6,8,10], G. Mandrile[11], G. Ferrero[11], A. Brusco[1,4]; [1]Dept. of Med. Sci., Univ. of Torino, Torino, Italy, [2]Inst. of Human Genetics and Ctr. for Molecular Med. Cologne, Univ. of Cologne, Cologne, Germany, [3]Dept. of Publ. Hlth.and Pediatrics, Univ. of Turin, Torino, Italy, [4]CittÃ della Salute e della Scienza Univ. Hosp., Torino, Italy, [5]Seaver Autism Ctr. for Res. and Treatment, Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Dept. of Psychiatry, Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]The Mindich Child Hlth.and Dev. Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [8]Friedman Brain Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [9]Icahn Sch. of Med. at Mount Sinai, New York, NY, [10]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [11]Dept. of Clinical and Biological Sci., Univ. of Torino, Torino, Italy
Disclosure Block: **C. Giovenino:** None.

In females, random X-chromosome inactivation (XCI) results in an approximate equal ratio of cells expressing maternal or paternal X chromosome genes. When XCI is non-random, this proportion can vary from mildly skewed XCI (e.g., 70:30) to completely skewed XCI (e.g., 0:100). Non-random XCI can be associated both to unaffected females (XCI favors the expression of the normal allele) or affected females. This latter case is explained by skewing towards a deleterious allele which reaches toxicity threshold or skewing away from a deleterious allele that decreases expression below lethality or an extreme phenotype.We explored the excess XCI in a clinically heterogeneous cohort of unsolved neurodevelopmental disease (NDD) cases - either the mothers of male patients or affected females - negative at FRAXA, array-CGH and Trio Whole Exome Sequencing (WES). We set up a fluorescent-PCR based screening for XCI on blood extracted DNA, amplifying three different microsatellites (AR, PCSK1N and SLITRK4 genes) after HhaI digestion. Currently, we analyzed 23 mothers of affected males and 45 female probands. We found a skewed XCI >80% in five of the 23 mothers (22%), and in ten of the 45 female patients (22%), far beyond the expected XCI in normal population (8%; p=0.0023, Chi square test). To search for a possible genetic mechanism determining XCI, we reanalyzed WES and clinical data. We possibly solved 5/15 cases (33%) whose causative gene was previously missed. In a boy, whose mother was 100% skewed, we found a previously missed causative deletion involving exons 3-4 in the ATRX gene (Mental retardation-hypotonic facies syndrome; MIM 309580). In two monozygotic twins (mother: 80% skewed XCI), we identified a possible causative c.-23C>A splicing variant in RLIM, associated with Tonne-Kalsheuer syndrome (MIM 300978), whose protein is involved in initiating XCI. In two brothers (mother: 100% skewed XCI), we identified a [c.1526C>T; p.(P509L)] missense variant in OTUD5, recently associated with X-linked multiple congenital anomalies-neurodevelopmental syndrome (MCAND; MIM 301056). Finally, we found a boy (mother: 100% skewed XCI) with a [c.2491C>T; p.(R831C)] missense variant in RBM10, associated with TARP syndrome (MIM 311900). Among affected females, we found the [c.1204G>A; p.(D402N)] likely pathogenic variant in KDM5C segregating in three other family members. Our work suggests to evaluate skewed XCI as an easy assay to further study cases without a diagnostic exome, to guide molecular re-evaluation for X-linked NDD. This approach may improve diagnostic yield and allow identifying new disorders with molecular mechanisms related to X chromosome inactivation.

PrgmNr 1022 - Understanding trajectories of neuronal differentiation underlying neurodevelopmental disorders using single-cell technologies

View session detail

Author Block: **D. Meistermann**[1], C. Roth[2], F. Iacoviello[1], P. PUIGDEVALL COSTA[3], H. Kilpinen[2,1,4]; [1]Helsinki Inst. of Life Sci. (HiLIFE), Helsinki, Finland, [2]Univ. Coll. London, London, United Kingdom, [3]UCL Inst. of Child Hlth., London, United Kingdom, [4]Wellcome Sanger Inst., Cambridge, United Kingdom
Disclosure Block: **D. Meistermann:** None.
The mechanisms behind severe neurodevelopmental disorders (NDDs) such as Kabuki syndrome (KS) are still elusive, despite their large impact in modern society. The main reason for this lack of knowledge is the complexity of the chain of events during development that lead to these conditions. Large genetic studies of severe NDDs have shown that they are commonly caused by highly penetrant de novo mutations that are frequently observed in genes encoding for chromatin modifying enzymes. In KS, the KDM6A and KMT2D genes have been shown to cause the majority of cases. To study how mutations in these genes might cause NDDs on the cellular level, we used induced pluripotent stem cells (iPSC) together with single-cell technologies to model the mutation effects on early neuronal development in vitro.
First, we differentiated iPSC lines obtained from 8 healthy individuals and 4 KS samples to neuronal precursor cells (NPC). We collected single-cell transcriptomic and chromatin accessibility profiles from these cells before and after differentiation and used the data to compare cell type composition and gene regulatory differences between patient and control cells. Second, to control for genetic background effects, we used CRISPR-Cas9 to create isogenic control iPSC lines as well as wild-type lines carrying engineered disease mutations. Third, to determine when KS-specific cellular phenotypes first appear, we profiled the cells daily over the course of NPC differentiation using immunocytochemistry and qPCR to identify a time point when KS-specific changes in cell identity and morphology start appearing. The experiment will then be repeated using additional cell lines and time points.
Collectively, these complementary data will be used to reconstruct the gene regulatory networks underlying Kabuki syndrome. With a better understanding of the molecular pathways and key developmental stages underlying this disorder, it may be possible to re-purpose existing drugs associated with these pathways to treat KS.

PrgmNr 1023 - *De novo* missense and truncating variants in *ZMYND8* result in a distinctive neurodevelopmental disorder

View session detail

Author Block: **K-R. Dias**[1,2], C. M. Carlston[3], L. E. R. Blok[4], L. De Hayr[5,6], U. Nawaz[7], C-A. Evans[1,8], P. Bayrak-Toydemir[9,10], S. Htun[11], Y. Zhu[8], A. Ma[12,13], S. Lynch[14], C. Moorwood[15], K. Stals[15], S. Ellard[15], M. N. Bainbridge[16], J. Friedman[16], J. G. Pappas[17], R. Rabin[17], C. B. Nowak[18], J. Douglas[18], T. E. Wilson[19], E. Torti[20], E. P. Kirk[8,21,22], J. R. Pinner[21,22], M. Edwards[23,24], A. Selvanathan[1,21,25], F. Montanari[26], C. Graziano[26], T. Pippucci[26], B. Dingmann[27], H. C. Mefford[28], T. Shimoji[29,30], T. Suzuki[31], K. Yamakawa[31], H. Streff[32], C. P. Schaaf[32,33], A. M. Slavotinek[11], I. Voineagu[34], J. C. Carey[35], M. F. Buckley[8], A. Schenck[4], R. J. Harvey[5,6], T. Roscioli[1,8,22]; [1]NeuroSci. Res. Australia, Sydney, Australia, [2]Prince of Wales Clinical Sch., Faculty of Med., Univ. of New South Wales, Sydney, Australia, [3]Div. of Genetics and Genomics, Boston Childrenâ□□s Hosp., Harvard Med. Sch., Boston, MA, [4]Dept. of Human Genetics, Donders Inst. for Brain, Cognition and Behaviour, Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [5]Sch. of Hlth.and Behavioural Sci., Univ. of the Sunshine Coast, Maroochydore, Australia, [6]Sunshine Coast Hlth.Inst., Birtinya, Australia, [7]Adelaide Med. Sch. and Robinson Res. Inst., The Univ. of Adelaide, Adelaide, Australia, [8]New South Wales Hlth.Pathology Randwick Genomics, Prince of Wales Hosp., Sydney, Australia, [9]Dept. of Pathology, Univ. of Utah, Salt Lake City, UT, [10]Molecular Genomics, ARUP Lab., Univ. of Utah Sch. of Med., Salt Lake City, UT, [11]Dept. of Pediatrics, Div. of Genetics, Univ. of California San Francisco, San Francisco, CA, [12]Children's Hosp. Westmead, Sydney Children's Hosp. Network, Sydney, Australia, [13]Discipline of Genomic Med., Sydney Med. Sch., Univ. of Sydney, Sydney, Australia, [14]Dept. of Clinical Genetics, Temple Street Children's Univ. Hosp., Dublin, Ireland, [15]Exeter Genomics Lab., Royal Devon and Exeter NHS Fndn. Trust, Exeter, United Kingdom, [16]Rady Children's Inst. of Genomic Med., Rady Children's Hosp., San Diego, CA, [17]Dept. of Pediatrics, Clinical Genetic Services, NYU Grossman Sch. of Med., New York, NY, [18]The Feingold Ctr. for Children, Boston Childrenâ□□s Hosp., Harvard Med. Sch., Boston, MA, [19]Dept. of Med. and Molecular Genetics, Indiana Univ. Sch. of Med., Riley Hosp. for Children at Indiana Univ. Hlth., Indianapolis, IN, [20]GeneDx, Gaithersburg, MD, [21]Sch. of Women's and Children's Hlth., Univ. of New South Wales, Sydney, Australia, [22]Ctr. for Clinical Genetics, Sydney Children's Hosp., Sydney, Australia, [23]Hunter Genetics, Hunter New England Hlth., New Lambton, Australia, [24]Sch. of Med., Western Sydney Univ., Penrith, Australia, [25]Discipline of Child and Adolescent Hlth., The Univ. of Sydney, Sydney, Australia, [26]U.O. Genetica Medica, IRCCS Azienda Ospedaliero-Univ.ria di Bologna, Bologna, Italy, [27]Med. Genetics Dept., Seattle Childrenâ□□s Hosp., Seattle, WA, [28]Ctr. for Pediatric Neurological Disease Res., St. Jude, Memphis, TN, [29]Dept. of Neurosurgery, Okinawa Pref. Nanbu Med. Ctr. and Childrenâ□□s Med. Ctr., Okinawa, Japan, [30]Okinawa Central Hosp., Okinawa, Japan, [31]Dept. of Neurodevelopmental Disorder Genetics, Inst. of Brain Sci., Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, [32]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [33]Heidelberg Univ., Inst. of Human Genetics, Heidelberg, Germany, [34]Sch. of Biotechnology and Biomolecular Sci., Univ. of New South Wales, Sydney, Australia, [35]Div. of Med. Genetics, Dept. of Pediatrics, Univ. of Utah Hlth., Salt Lake City, UT

Disclosure Block: **K. Dias:** None.

ZMYND8 encodes a protein that serves as a central interactive hub for coordinating critical roles in transcription regulation, chromatin remodeling, regulation of super-enhancers, DNA damage response and tumor suppression. Through exome sequencing and an international collaboration, we assembled genotype and phenotype data for 15 affected, unrelated individuals with *de novo* variants in *ZMYND8*, of which 12 are missense, three are truncating and one is recurrent. This novel neurodevelopmental disorder is characterized by moderate to profound intellectual disability, non-familial facial features, cardiovascular, ophthalmologic and minor skeletal anomalies. Molecular modelling and yeast two-hybrid assays suggest that missense variants in the PWWP domain of ZMYND8 abolish interactions

with Drebrin, which translocates the ZMYND8 protein from the nucleus to the cytoplasm of neurons, while missense variants in the MYND domain abolish interactions with GATAD2A, a key component in the NuRD complex. Based on available human brain transcriptomic data, we demonstrate that ZMYND8 is broadly expressed across all cell types in all brain regions and shows higher expression in the early stages of brain development compared to the postnatal period. Neuronal knockdown of the *Drosophila ZMYND8* ortholog results in decreased habituation learning, consistent with a role of the gene in cognitive function. In conclusion, we present genomic and functional evidence for disruption of *ZMYND8* as a novel etiology of syndromic intellectual disability.

PrgmNr 1024 - Extremely rare variants in *EIF4A2* are associated with a neurodevelopmental disorder characterized by hypotonia, intellectual disability and epilepsy

View session detail

Author Block: **M. Paul**[1,2], A. Duncan[3], C. A. Genetti[4,5,6], M. Pinelli[7,8], N. Brunetti-Pierri[9,8], A. Garza-Flores[10], R. Saneto[11], E. Agolini[12], U. BlÃ¼mlein[13], R. Abou Jamra[14], R. Carapito[15], I. Bertrand[16], S. Bahram[17], A. Ritter[18], K. Izumi[19], B. Ben Zeev[20], H. Pan[2,21], H-T. Chao[1,2,22,23,24], P. B. Agrawal[5,6,4]; [1]Dept. of Pediatrics, Div. of Neurology and Dev.al NeuroSci., Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst., Texas Childrenâ□□s Hosp., Houston, TX, [3]Boston Childrenâ□□s Hosp. and Harvard Med. Sch., Boston, MA, [4]Div. of Newborn Med., Boston Childrenâ□□s Hosp. and Harvard Med. Sch., Boston, MA, [5]Div. of Genetics and Genomics, Boston Childrenâ□□s Hosp. and Harvard Med. Sch., Boston, MA, [6]The Manton Ctr. for Orphan Disease Res., Boston Childrenâ□□s Hosp. and Harvard Med. Sch., Boston, MA, [7]Pozzuoli, Italy, [8]Univ. of Naples 'Federico II', Pozzuoli, Naples, Italy, [9]Telethon Inst Genetics & Med., Pozzuoli, Italy, [10]Cook Childrenâ□□s Hosp., Fort Worth, TX, [11]Seattle Children's Hosp., Seattle, WA, [12]Ospedale Pediatrico Bambino GesÃ¹, Roma, Italy, [13]Carl-Thiem-Klinikum, Cottbus, Germany, [14]Inst. of Human Genetics, Univ. of Leipzig Med. Ctr., Leipzig, Germany, [15]INSERM / Univ. of Strasbourg, Strasbourg, France, [16]Ctr. Hosp.ier Univ.ire de Nantes, Nantes, France, [17]Faculte de Medecine, Strasbourg, France, [18]Div. of Human Genetics and Metabolism, Dept. of Pediatrics, The Children's Hosp. of Philadelphia, Philadelphia, PA, [19]Children's Hosp. of Philadelphia, Philadelphia, PA, [20]Pediatric Neurology Dept., The Edmond and Lilly Safra Pediatric Hosp., Sheba Med. Ctr., Tel HaShomer, Israel, [21]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [22]Dept. of Molecular and Human Genetics, Dept. of NeuroSci., Baylor Coll. of Med., Houston, TX, [23]Texas Childrenâ□□s Hosp., Houston, TX, [24]McNair Med. Inst., The Robert and Janice McNair Fndn., Houston, TX

Disclosure Block: **M. Paul:** None.

Eukaryotic Initiation Factor-4A2 encodes EIF4A2, an ATP-dependent RNA helicase subunit of the eIF4F complex, which recognizes the 5â□□cap structure of mRNAs and is required for mRNA binding to the ribosome. The fruit fly homolog *eIF4A* mediates the negative regulation of Decapentaplegic (Dpp) signaling. In the fly, Dpp-signaling regulates embryo patterning, eye and wing morphogenesis, and stem cell identity determination. The vertebrate homolog of Dpp, TGF-Î²/BMP, is a key regulator of neuronal differentiation, development, and function. Dysregulation of TGF-Î²/BMP signaling is associated with various neurological disorders. Prior fly studies revealed that both gain and loss of function (GOF and LOF) *eIF4A* alleles modulate the rough eye and wing serration phenotypes associated with Dpp GOF and LOF, respectively. Despite the role of EIF4A2 homologs in key developmental processes, human disease-causing variants have not previously been identified. Here, we report eleven individuals with *extremely rare* variants in *EIF4A2* who all present with global developmental delay or intellectual disabilities, significant hypotonia, and epilepsy in most cases. To determine the pathogenicity of the *EIF4A2* variants *in vivo*, we generated transgenic flies expressing human *EIF4A2* wild-type (WT) and variants with a C-terminal HA tag for four *de novo* variants. We used the GAL4-UAS system to selectively express human EIF4A2 in fly neurons, wing, or eye. First, we conducted climbing assays to determine the impact of neuronal expression of EIF4A2 p.L344F, p.G364E, and p.T243I. We found that expression of these variants resulted in motor defects. Second, we found that the wing specific expression of EIF4A2 p.T216I caused wing serration, which is consistent with loss of Dpp signaling. This finding indicates that p.T216I is a GOF variant. Third, we found that the eye specific overexpression of EIF4A2 p.L344F exacerbates the rough eye phenotypes associated with Dpp GOF, suggesting that p.L344F is a LOF variant. Together, these findings reveal that these *de novo EIF4A2* variants are pathogenic and alter fruit fly development in a dominant

manner through either LOF or GOF mechanisms. Our work establishes a role for EIF4A2 dysfunction in human neurodevelopmental disorders.

PrgmNr 1025 - Autism spectrum disorder trios from consanguineous populations are enriched for rare biallelic variants

View session detail

Author Block: **R. S. Harripaul**; Ctr. for Addiction and Mental Hlth., Ajax, ON, Canada
Disclosure Block: **R.S. Harripaul:** None.
Autism spectrum disorder (ASD) is a severe neurodevelopmental disorder that affects about 1 in 55 children worldwide and imposes enormous economic and socioemotional burdens on families and communities. Genetic studies of ASD have identified de novo copy number variants (CNVs) and point mutations that contribute significantly to the genetic architecture of ASD, but the majority of these studies were conducted in outbred populations, which are not ideal for detecting autosomal recessive (AR) inheritance. However, several studies have investigated ASD genetics in consanguineous populations and point towards AR as an under-appreciated source of ASD variants. Here, we used trio whole-exome sequencing (WES) to look for rare variants for ASD in 115 proband-mother-father trios from populations with high rates of consanguinity, namely Pakistan, Iran, and Saudi Arabia. In total, we report 93 candidate sequence variants, with 57% biallelic, 24% dominant/de novo, and the rest X-linked. 52% of the variants were loss of function (LoF) or putative LoF (splice site, stop loss) and 47% non-synonymous. Our analysis indicates enrichment of previously identified and candidate AR genes. These include variants in genes previously reported for AR ASD and/or intellectual disability (ID), such as AGA, ASL, ASPA, BTN3A2, CC2D1A, DEAF1, HTRA2, KIF16B, LINS1, MADD, MED25, MTHFR, RSRC1, SHH, TECPR2, VPS13B, ZNF335, and 36 previously unreported candidates, including LoF variants in genes such as DAGLA, EFCAB8, ENPP6, FAXDC2, ILDR2, PKD1L1, SCN10A, and SLC36A1. We also identified likely causative CNVs in N individuals, and candidate biallelic exonic loss CNVs in four trios, implicating the genes SEMG1, DNAH7, DHRS4L2, and SIRPB1 Gene expression and cell-typing analysis indicated that the cortical plate, subplate and ventricular and excitatory and inhibitory neurons in critical learning and developmental regions such as the hippocampus and cerebral cortex were significantly enriched.

PrgmNr 1028 - Long read sequencing to improve diagnostic rate in critically ill patients

View session detail

Author Block: **M. N. Bainbridge**[1], A. Wenger[2], S. Batalov[3], J. Stone[2], P. Baybayan[2], C. Lambert[2]; [1]Rady Children's Inst. for Genomic Med., San Diego, CA, [2]Pacific BioSci.s, Menlo Park, CA, [3]Rady Childrens Inst., Genomic Med., San Diego, CA
Disclosure Block: **M.N. Bainbridge:** None.
Despite advances in sequencing and analytics, diagnostic rates for critically ill children remain in the 30-40% range. We selected a cohort of 22 patients, who remained undiagnosed after analysis both clinically and through a research pipeline of their 40x WGS short read sequenced genomes. Phenotypes of patients ranged from epilepsy, syndromic autism, complex undetermined syndromes and sudden infant death. We also selected three patients with known genetic diagnosis that were difficult to identify with short reads. We sequenced this cohort of patients to 10-30X depth of coverage using Pacific Biosciences HiFi long-read technology to assess whether there was an increase in diagnostic rate.

Of the three controls, long read successfully identified a pathogenic, hemizygous frameshift mutation in IKBKG. Due to high sequence homology this variant was not callable with short reads and patient was only diagnosed from a gene specific assay. Long reads were also able to detect a 60Mbp inversion which disrupted EYA1, the cause of autosomal dominant Branchiootic syndrome 1 [MIM:602588]. Finally, long reads identified a 70Kbp contig that established an apparent translocation of between chromosomes 22 and 11. This patient was diagnosed with a derivative, supernumerary chromosome of 22 and 11 (Emanuel syndrome).

Of the undiagnosed patients, long reads identified an average of 52 structural variants that occurred in known disease genes per patient; this included a large deletion of FBN2. Intriguingly, long reads were also able to detect an average of 2,184 small (Long read sequencing can identify numerous variants, both small and structural that are not readily detectable by short reads. The number of small, coding variants in disease genes missed by short reads is of particular concern and may impact the overall diagnostic rate for critically ill children.

PrgmNr 1029 - Establishing nonhuman primate models of human disease using a publicly available rhesus macaque genomic variant database

View session detail

Author Block: **S. Peterson**[1], A. Lewis[1], E. Vallender[2], A. Ericsen[3], K. Ray[1], B. N. Bimber[1], B. Ferguson[1]; [1]Oregon Hlth.& Sci. Univ., Portland, OR, [2]Univ Mississippi Med Ctr, Jackson, MS, [3]Emory Univ., Atlanta, GA

Disclosure Block: **S. Peterson:** None.

Rhesus macaques serve a vital role as preclinical models of disease due to their high level of genetic and physiological homology with humans. The discovery of new genetic models has been greatly assisted by publically available genomic data. The Macaque Genotype And Phenotype Resource (mGAP; https://mgap.ohsu.edu/) catalogs whole genome and whole exome sequence data from rhesus macaques from the Oregon National Primate Research Center (ONPRC) as well as other National Primate Research Centers and external collaborators. The current release (mGAP 2.1) includes genetic data from 2,329 rhesus macaques mapped onto the latest macaque genome (Mmul10) along with a LiftOver track to visualize the data on the human genome. Variants are extensively annotated with relevant information including the location relative to genes, population frequency, sequence conservation, and scores from various variant effect prediction tools (CADD, SnpEff, PolyPhen2) or those variants which identically match pathogenic or suspected pathogenic human ClinVar variants. Variants predicted to be damaging to over 2,700 OMIM disease/phenotypes have been identified to date. Critically, the variant information in this database includes subject-level genotype data allowing the resource to be leveraged along with extensive health information records and pedigree information for each monkey. This has allowed for the identification of spontaneous non-human primate models of disease either through the matching of observed disease phenotypes with a causative genetic variant or by facilitating pedigree screens of individuals who carry specific alleles of interest for specific signs of disease. These approaches have been successful in establishing numerous models of human disease, including epidermolysis bullosa, Bardet-Biedl Syndrome, oculocutaneous albinism, hemochromatosis, cobblestone lissencephaly, osteochondrodysplasia and dyshormonogenetic goiter. As we rapidly expand this list, these verified models are being used to further disease investigation through biomarker discovery, gene therapy approaches and the study of alternative metabolic pathways. Funded by NIH R24 OD021324

PrgmNr 1030 - Utilization of Dual-Label Optical Genome Mapping for genetic/epigenetic diagnosis

View session detail

Author Block: **S. Bhattacharya**[1], H. Sharim[2], Y. Ebenstein[2], Y-W. Chen[1], S. I. Berger[3], E. J. Vilain[3], H. Barseghyan[4]; [1]Children's Natl. Hosp., Washington, DC, [2]Tel Aviv Univ., Tel Aviv, Israel, [3]Children's Natl. Med. Ctr., Washington, DC, [4]Childrens Natl. Hosp., Washington, DC
Disclosure Block: **S. Bhattacharya:** None.
Short-read exome/genome sequencing (SRS) and chromosomal microarrays (CMA) have helped increase diagnostic rates across many genetic disorders. However, despite this success, about half of the cases remain undiagnosed. Due to the methodological limitations of both technologies (SRS, CMA) they fail to sensitively identify structural variants or balanced rearrangements, respectively. Additionally, both technologies have limitations in assessment of epigenetic changes. Short-read based bisulfite sequencing or methylation arrays do not provide long-range haplotype specific methylation states, rather the detected signals are averaged for individual genomic positions. These limitations can be alleviated with a novel dual-label optical genome mapping (DL-OGM) technology for detection of both genetic and epigenetic changes in one assay over long stretches of single DNA molecules and phased haplotypes. The method relies on differential labeling of high molecular weight DNA. First, long DNA molecules are nicked with BspQI endonuclease and labeled with red fluorescent nucleotides. Second, the same DNA molecules undergo treatment with M.TaqI methyltransferase that attaches green fluorescent cofactor onto non-methylated CpGs in ATCG sequences throughout the genome. Third, the pattern of fluorescent labels is captured in nanochannel arrays for de novo genome assembly, variant calling and quantification of epigenetic marks. Additionally, separate bioinformatics pipelines have been developed to annotate and filter the variants and quantify the methylation marks, producing robust genomic and epigenomic results. Here, we show the ability of DL-OGM to detect large copy number variants and methylation levels for Facioscapulohumeral muscular dystrophy (FSHD) and Beckwith-Wiedemann syndrome (BWS). We successfully identified the molecular diagnosis (constriction of D4Z4 array and associated hypomethylation) in FSHD case/control samples in the sub-telomeric region of chromosome 4q35. Additionally, we tested the method for a case diagnosed with BWS, where DL-OGM identified a duplication in the paternally inherited allele carrying epigenetic states resulting in the syndrome. DL-OGM technology offers substantial advantages over the current clinical diagnostic practices for specific disorders tested here (FSHD, BWS) and can be applied to other types of disorders such as CHARGE syndrome.

PrgmNr 1031 - 2D and 3D human neuronal cell models of *CHD2*-associated epilepsy

View session detail

Author Block: **K-M. J. Lamar**[1], J. Gunti[1], A. Tidball[2], J. Z. Cheng[1], J. Parent[2], G. L. Carvill[1];
[1]Northwestern Univ., Chicago, IL, [2]Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **K.J. Lamar:** None.
Pathogenic variants in *CHD2* are associated with developmental epileptic encephalopathy (DEE) in humans. The majority of these variants are frameshifts or deletions and all are heterozygous, having arisen *de novo*, indicating that haploinsufficiency is the mode of pathogenesis.
Chromodomain helicase DNA binding protein 2 (CHD2) is a chromatin remodeler that is associated with sites of active transcription.To better understand the role of CHD2 in human DEE, we utilized the CRIPSR/Cas9 system to create heterozygous disruptions of *CHD2* in human induced pluripotent stem cells (iPSCs). To examine the effect of *CHD2* disruption in neuronal development, we differentiated these iPSC lines into neuronal progenitor cells (NPCs) and cerebral organoids (COs). We performed RNA sequencing on $CHD2^{+/-}$ NPCs and isogenic wild-type (WT) controls and found that $CHD2^{+/-}$ NPCs exhibited an upregulation of genes involved in neuronal differentiation and a downregulation of genes involved in proliferation. CHD2 disruption in NPCs also lead to misregulation of other epilepsy genes. We performed single cell RNA sequencing (scRNAseq) in Day 48 COs from two WT lines and two $CHD2^{+/-}$ lines to determine whether *CHD2* disruption in distinct neuronal subtypes causes misregulation of genes important for human neurogenesis. Analysis of scRNAseq from COs revealed that the organoids form distinct neuronal subtypes, including radial glia, outer radial glia, intermediate progenitor cells, excitatory neurons, and interneuron progenitors. Differential expression analyses are ongoing to determine in which neuronal cell types *CHD2* loss has the biggest effects on gene regulation.
Overall, our results suggest that *CHD2* haploinsufficiency results in a cell proliferation defect which likely contributes to premature neuronal differentiation. In the future we aim to identify CHD2 targets and epigenetic changes during development. Ultimately, we seek to understand the mechanisms leading to gene expression changes when *CHD2* is disrupted and the role of CHD2 in regulating other epilepsy genes. The role of chromatin remodeling and control of gene expression is largely unexplored in epilepsy, but it is a dynamic process with many therapeutic targets and presents a novel opportunity to develop new treatments for patients.

PrgmNr 1032 - Long-read Sequencing Identifies Missing Disease-Causing Variants and Resolves Complex Structural Variants

View session detail

Author Block: **D. Miller**[1], M. Galey[1], E. E. Eichler[2]; [1]Univ. of Washington, Seattle, WA, [2]Univ of Washington, Seattle, WA
Disclosure Block: **D. Miller:** None.
Despite the widespread use of clinical genetic testing, a substantial fraction of patients with suspected genetic diseases remain undiagnosed. In some cases, clinical testing reveals structural differences that are challenging to fully evaluate, while in other cases, a suspected gene cannot be adequately evaluated due to technical limitations of short reads. Long-read sequencing of native DNA overcomes many of the limitations of existing clinical testing, but is expensive to implement, it has not yet been evaluated for clinical use. We performed targeted long-read sequencing (T-LRS) on an Oxford Nanopore platform using Adaptive Sampling, which allowed us to computationally select specific regions of the genome for sequencing in real-time. We targeted up to 150 Mbp of sequence per experiment, resulting in an average of 20-50 fold sequence coverageÂ¬â□□a significant increase above background. After base calling and alignment, we identified single nucleotide variants and structural variants and found that our method had 100% concordance with variants identified on clinical testing. For cases in which T-LRS could not resolve complex structural variants that lie within highly repetitive regions we performed whole-genome long-read sequencing (WG-LRS) on the Nanopore platform. Pathogenic and likely pathogenic missing disease-causing variants were identified in individuals for whom traditional genetic testing revealed only a single pathogenic variant for a known autosomal recessive disorder or was nondiagnostic for a suspected autosomal dominant or X-linked disorder. For example, T-LRS of an individual with biochemically confirmed Lesch-Nyhan syndrome revealed a 17-Mb inversion of the X chromosome that bisects HPRT1 within an intron that was not identified by targeted clinical sequencing. T-LRS was also used to phase known variants, fully sequence resolve triplet repeat expansions as well as their methylation state and resolve complex copy number changes discovering novel rearrangement breakpoints of functional and clinical relevance. Assembly and analysis of WG-LRS data resolved complex structural changes and demonstrates limitation of T-LRS in large repetitive regions of the genome. We show that T-LRS can be used as a single, cost-effective genetic test to accurately resolve pathogenic structural variants, methylation status and identify single-nucleotide variants not identified by other technologies. Overall, long-read sequencing will facilitate gene discovery, discover missing genetic variants and provide clarity in difficult-to-access regions of the human genome, improving the rate of clinical diagnosis.

PrgmNr 1033 - Using CRISPR-based saturation genome editing to improve the diagnosis of neuro-developmental disorders

View session detail

Author Block: **E. Radford**[1,2], H. Tan[2], M. Andersson[2], H. Ironfield[2], J. Stephenson[3], S. Gerety[2], M. E. Hurles[2]; [1]Univ. of Cambridge, Cambridge, United Kingdom, [2]Wellcome Sanger Inst., Cambridge, United Kingdom, [3]EMBL-EBI, Cambridge, United Kingdom
Disclosure Block: **E. Radford:** None.
Genetic sequencing is a powerful and widely used diagnostic tool. However, correctly interpreting candidate genetic variants remains a major challenge, and the number of variants of uncertain significance (VUS) is growing exponentially. Predictive testing and some targeted therapies cannot be offered where VUS are identified. Conventional strategies to resolve VUS rely on the accumulation of clinical data and variant effect prediction algorithms. Clinical data accumulate too slowly to be useful to families with rare neurodevelopmental disorders. Variant effect prediction algorithms are confounded by circularity and error propagation and are often discordant. New approaches are urgently needed.
Saturation genome editing (SGE) utilises CRISPR-Cas9 technologies to specifically engineer thousands of genetic variants into the endogenous genetic locus of a pool of cells and directly test their functional impact *in vitro*. The use of a haploid human cell line (HAP1) simplifies the design and interpretation of genetic screens. As proof-of-principle we have applied SGE to *DDX3X*, an X-linked RNA helicase of the DEAD-box protein family which is essential in HAP1. Heterozygous *DDX3X* loss-of-function mutations are one of the commonest genetic causes of female intellectual disability (Snijders-Blok *et. al*, 2015). We have functionally characterised every possible *DDX3X* single nucleotide variant and codon deletion, the majority of which have not previously been observed in the human population. SGE correctly identifies synonymous variants as functional, while the majority of loss-of-function variants are non-functional - these variants abrogate the expression or function of *DDX3X*. SGE is therefore a sensitive and specific test of loss of *DDX3X* function. Missense variants show a bimodal distribution, allowing us to categorise which variants render the protein non-functional. Missense variants within the helicase domain of *DDX3X* (exons 12-15), are more likely to render the protein non-functional. Strikingly, in the final and penultimate exons, many nonsense variants remain functional, consistent with these regions escaping nonsense-mediated decay. The majority of variants that have been reported as likely/pathogenic in a clinical context are identified as non-functional in our assay, while those variants observed in population studies of healthy individuals (UK Biobank and GnomAD) are identified as functional. Together these data suggest that SGE in HAP1 cells is a valuable approach to functionally assess developmental disorder genetic variants at scale. We are now scaling up this approach to assess many developmental disorder-associated genes.

PrgmNr 1036 - Single cell â□□whole embryoâ□□ phenotyping of developmental disorders

View session detail

Author Block: **J. Henck**[1], X. Huang[2], C. Qui[2], U. Kornak[3], D. Ibrahim[1], A. Despang[1], C. Trapnell[2], L. Saunders[2], I. Kurth[4], N. Haag[4], M. Robson[1], A. Ringel[1], W-L. Chan[5], V. Kalscheuer[1], L. Wittler[1], D. Beier[6], D. E. Dickel[7], A. Visel[7], L. A. Pennacchio[8], J. Cao[9], J. Shendure[2], M. Spielmann[1]; [1]Max Planck Inst. for Molecular Genetics, Berlin, Germany, [2]Univ. of Washington, Seattle, WA, [3]Univ. Med. Ctr. GÃ¶ttingen, GÃ¶ttingen, Germany, [4]RWTH Aachen Univ., Aachen, Germany, [5]CharitÃ©-Univ.smedizin Berlin, Berlin, Germany, [6]Univ. of Washington Sch. of Med., Seattle, WA, [7]Lawrence Berkeley Natl. Lab., Berkeley, CA, [8]Lawrence Berkeley Natl. Lab, Berkeley, CA, [9]Rockefeller Univ., New York, NY
Disclosure Block: **J. Henck:** None.
Mouse models represent a critical tool to study human diseases, particularly developmental disorders. A fundamental challenge to study pleiotropic developmental disorders *in vivo* is the lack of current technologies with sufficient throughput and resolution to obtain a global view of the molecular states and trajectories of a rapidly diversifying and expanding number of cell types during embryogenesis. We set out to establish single cell RNA sequencing at the â□□whole embryoâ□□ scale as a tool for unbiased phenotyping of mouse genetic models. In a single experiment, we applied combinatorial indexing based sc-RNA-seq to profile 103 embryos of 22 different mouse mutants and 4 different wildtype backgrounds at embryonic stage E13.5. Towards evaluating the sensitivity of this approach, the severity of the mouse mutants investigated range from established multisystem disorders to single enhancer knockouts. The resulting mouse mutant cell atlas (MMCA) consists of over 1.9 million single cell RNA-seq profiles. We developed an analytical framework for molecular phenotyping of cell type and trajectory composition changes, gene expression alterations and developmental phenotypes. We identify mutant and strain specific cell type changes, compare phenotyping of gain and loss of function mutants, and characterize deletions of topological associating domains boundaries. Intriguingly, even amongst these 22 mutants, some changes are shared by distinct models, suggesting that the molecular and cellular correlates of the recurring â□□componentsâ□□ of pleiotropic phenotypes might be recovered by further scaling of this approach. Overall, our findings show that â□□whole embryoâ□□ single cell profiling represents a powerful tool to systematically investigate developmental disorders at unprecedented resolution.

PrgmNr 1037 - Identifying systematic variation at the single-cell level by leveraging low-resolution population-level data

View session detail

Author Block: **E. Rahmani**[1], E. Halperin[2], M. I. Jordan[1], N. Yosef[1]; [1]UC Berkeley, Berkeley, CA, [2]Univ California los Angeles, Los Angeles, CA
Disclosure Block: **E. Rahmani:** None.
INTRODUCTION: Existing single-cell (SC) datasets are limited by their number of donors (individuals). As a result, most of the current research in SC genomics focuses on studying processes that are broadly conserved across individuals, such as differentiation and tissue development. While studying such biology from a limited number of donors is possible due to the expected high consistency across donors, advancing our understanding of heterogeneous conditions that demonstrate molecular variation and therefore inconsistencies across individuals requires population-level data.
OBJECTIVE: Our goal was to develop an approach to integrate information from low-resolution yet large bulk data into the analysis of small yet high-resolution SC data, and to demonstrate its promise by detecting systematic singe-cell variation in gene-gene interactions with a heterogeneous condition.
METHODS: We developed â□□kernel of integrated single cellsâ□□ (Keris). By inferring cell-type-specific moments (means, variances, and gene-gene interactions) and their variation with conditions in large tissue-level bulk data representing a population, Keris allows us to generate testable hypotheses at the SC level. Particularly, population-based hypotheses generated by Keris can be interpreted in terms of single-cell variation and hence can be casted as hypotheses on populations of cells; this enables validation and further exploration using SC data from just a handful of donors.
RESULTS: We applied Keris to bulk PBMC data from two age groups (n=745; all 70 y/o) for detecting systematic cell-type-specific variation in gene-gene interactions with cellular senescence; we verified that achieving this goal in SC data alone would require a large number of donors, given the heterogeneity in aging (even within the same age group). Using PBMC from 4 donors (38, 70, 84, and 90 y/o), we first verified concordance between the Keris-predicted up-regulated interactions and expression of *CDKN2A*, a marker of cellular senescence. Then, we composed a cell-level senescence score by integrating expression across the Keris-derived gene-gene interactions and successfully used it to detect variation in cell populations across the trajectory of the senescence score; for example, we identified clear stratification of sub-populations of CD16 monocytes and natural killer cells according to our cellular senescence scores in all 4 donors.
CONCLUSIONS: Applying Keris to large bulk data identifies consistent single-cell variation in gene-gene interactions, thus enabling effective downstream analysis of heterogenous conditions using SC data without the need for a large number of donors.

PrgmNr 1038 - Discovery of target genes and pathways of blood trait loci using pooled CRISPR screens and single cell RNA sequencing

View session detail

Author Block: **J. A. Morris**[1,2], Z. Daniloski[1,2], J. Domingo[1], T. Barry[3], M. Ziosi[1], D. A. Glinos[1,4], S. Hao[1], E. P. Mimitou[1], P. Smibert[1], K. Roeder[3], E. Katsevich[5], T. Lappalainen[1,4], N. E. Sanjana[1,2]; [1]New York Genome Ctr., New York, NY, [2]New York Univ., New York, NY, [3]Carnegie Mellon Univ., Pittsburgh, PA, [4]Columbia Univ., New York, NY, [5]Univ. of Pennsylvania, Philadelphia, PA

Disclosure Block: **J.A. Morris:** None.

The majority of variants associated with complex traits and common diseases identified by genome-wide association studies (GWAS) map to noncoding regions of the genome with unknown regulatory effects in *cis* and *trans*. To discover target genes of noncoding variants, we developed an approach by leveraging biobank-scale GWAS data, massively parallel CRISPR screens and single cell sequencing. To identify candidate cis-regulatory elements (CREs), we first performed statistical fine-mapping with FINEMAP to identify putative causal SNPs from 29 quantitative blood trait GWAS in the UK Biobank. We then selected a subset of 88 noncoding SNPs from 56 loci that overlap DNase, ATAC-seq and H3K27ac ChIP-seq peaks in human erythrocyte progenitor cells (K562) for targeting with multiple CRISPR guide RNAs (gRNAs). We silenced these loci in a pooled manner in K562 cells that constitutively express a second-generation CRISPR repressor (MeCP2-dCas9-KRAB). Single cell RNA sequencing of 9,343 single cells, with an average of 10 gRNAs per cell, allowed us to test for differential expression induced by each gRNA. We identified significant target genes in *cis* (5% FDR) for nearly half of the 88 SNPs. For example, gRNAs targeting *rs4526602*, a candidate causal GWAS SNP for a monocyte percentage locus, were significantly associated only with *CD52* expression and not to two closer genes *AIM1L* or *UBXN11*.

Next, we tested if any gRNAs with *cis* effects also had downstream effects on the transcriptome and identified three SNPs with significant target genes in *trans*. For example, another lead SNP in a monocyte GWAS (*rs524137*) that mapped to an intergenic CRE for transcription factor *GFI1B* (-25 kb from its TSS) was also associated with 568 other genes in trans (1% FDR). The TSSs of these target genes were enriched for having *GFI1B* ChIP-seq peaks in K562 cells (odds ratio = 2.4, $p = 7.4x10^{-23}$] and had a replication rate of 56% in an independent data set. We further showed that the target genes form a coexpression network with multiple modules reflecting both direct and indirect targets of *GFI1B* that are expressed in human bone marrow progenitor cell types.

In summary, we have developed a high-throughput approach integrating genome engineering, single cell sequencing and computational methods to identify causal GWAS SNPs and connect them to their target genes. We identified target genes for blood trait loci and identified *trans* target gene networks. This methodology will enable massively parallel assays to catalog the target genes of human noncoding variants in both *cis* and *trans*.

PrgmNr 1039 - SIMBA-single cell embedding along with features

View session detail

Author Block: **H. Chen**[1], J. Ryu[1], M. Vinyard[1], A. Lerer[2], L. Pinello[1]; [1]Massachusetts Gen. Hosp./ Harvard Med. Sch., Boston, MA, [2]Facebook AI Res., Boston, MA
Disclosure Block: **H. Chen:** None.
Recent advances in single cell omics technologies enable the individual or joint profiling of cellular measurements including gene expression, epigenetic features, chromatin structure and DNA sequences.
Currently, single-cell analysis pipelines are mostly cluster-centric, i.e., they first construct cellular states based on clustering approaches and then extract their defining features. This characterization discretizes the continuous nature of cellular heterogeneity, highlighting only those features important to non-overlapping cell groups.
To address this shortcoming, we developed SIMBA, a method to embed cells along with their defining features such as gene expression, transcription factor binding sequences and chromatin accessibility peaks into the same latent space. In SIMBA, cells and features from different modalities are encoded as nodes and relations between them as edges in a single graph. These relations encode our knowledge on the different aspects of gene regulation. For example, if a cell expresses a gene, an edge between the cell and gene nodes is created. Similarly, an edge can encode a chromatin accessible peak in a cell, a TF binding in a peak, the DNA sequence of a peak, etc.
Low-dimensional representations for each of the nodes in this graph are then computed using an unsupervised graph embedding method. The resulting joint embedding of cells and features in a common latent space provides a non-cluster-centric and clustering-free way to study intrinsic relations between biological entities. Moreover, our efficient graph embedding procedure makes SIMBA easily scalable to millions of cells. Importantly we show that in addition to single-modal analyses (e.g., scRNA-seq or scATAC-seq), several current single-cell analysis tasks including batch correction, data integration and multimodal analysis can be naturally incorporated in this general framework.
We have extensively tested SIMBA on seven scRNA-seq datasets, five scATAC-seq datasets, and three dual-omics datasets. In each analysis, SIMBA performs comparably to or better than current state-of-the-art methods specifically developed for each task.
In summary, our study proposes a novel single-cell embedding method SIMBA. Our joint embedding of cells and features provides a new non-cluster-centric and clustering-free framework to study single-cell heterogeneity and mechanisms of gene regulation. SIMBA is available as a comprehensive open-source Python package at https://github.com/pinellolab/simba.

PrgmNr 1040 - Deciphering the variation of gene regulation among individuals via integrative single cell multi-omics analysis

View session detail

Author Block: **J. Wang**[1], X. Cheng[1], Q. Liang[1], M. Wang[1], M. DeAngelis[2], Y. Li[1], R. Chen[1]; [1]Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Dept. of Ophthalmology, Univ. at Buffalo The State Univ. of New York, New York, NY
Disclosure Block: **J. Wang:** None.
Gene regulation variation between individuals is an essential molecular phenotype underpinning human traits and diseases, and genes are regulated in a cell type specific manner. The recent advances in single-cell omics technologies offer unprecedented opportunities to evaluate the effect of genetic variants on gene regulation within each cell type in vivo. To probe the underlying drivers of gene regulation variation, retinae from 20 human donors were profiled with single nucleus RNA-sequencing (snRNA-seq) and single nucleus ATAC-sequencing (snATAC-seq). The 191,818 nuclei from snRNA-seq and 245,835 nuclei from snATAC-seq were assigned to one of ten major retinal cell types. The genes with variable expression and genomic regions with variable chromatin accessibility among individuals are identified for each cell type. In addition, whole genome sequencing was performed to obtain the genotype for each donor. By systematically analyzing the association between genetic variants, gene expression, and chromatin accessibility in each cell type, a list of single-cell expression QTLs (sc-eQTLs), single-cell chromatin accessibility QTLs (sc-caQTLs), single-cell allelic specific expression (sc-ASE), and single-cell allelic specific chromatin accessibility (sc-ASCA) were identified. Strikingly, a significant proportion of these associations are cell type specific, highlighting the power of single cell analyses. By combining sc-eQTL, sc-caQTL, sc-ASE, sc-ASCA, snATAC-peaks, and single cell gene expression, we identified a set of cis-elements that are likely to directly contribute to cell-type specific variation in gene expression. Furthermore, we uncovered the genetic variants that are likely to affect the activity of gene cis regulatory elements with cell-type specific effects. Finally, by integration of results from the single-cell datasets and bulk eQTL and GWAS, we fine-mapped causal variants and targeted genes in cell type resolution.

PrgmNr 1041 - scATAC.Explorer: A Single-Cell ATAC-Seq Database and Search Tool

View session detail

Author Block: **A. Gibson-Khademi**[1], E. Christensen[2], P. Shooshtari[3]; [1]Pathology and Lab. Med., Univ. of Western Ontario, London, ON, Canada, [2]Computer Sci., Univ. of Western Ontario, London, ON, Canada, [3]Children Hlth.Res. Inst., London, ON, Canada
Disclosure Block: **A. Gibson-Khademi:** None.
Single Cell Assay for Transposase-Accessible Chromatin using Sequencing (scATAC-seq) is a relatively new technology that allows for identifying regions of accessible chromatin at a single cell resolution, giving insights to epigenetic differences in tissues, cell types, and diseases. Although several scATAC-seq datasets have been generated recently, we still lack a comprehensive database of scATAC-seq that contains a wide-range of datasets in a consistent format. We have addressed this problem by creating an R accessible database and search tool that can query a curated collection of currently available scATAC-seq datasets, and return all datasets and metadata in a consistent format.
We gathered scATAC-seq datasets and associated metadata from NCBIâ□□s Gene Expression Omnibus (GEO) and research lab websites. Several different data formats were found. Datasets were then formatted into a consistent peak-by-cell matrix format using R. Cell type annotations or cell clustering assignments were also collected if available. We then developed an R package named scATAC.Explorer that allows for users to query for relevant datasets using parameters such as cell type, disease, and other dataset metadata. Datasets retrieved from scATAC.Explorer are available as SingleCellExperiment R objects for ease of use with Bioconductor R packages.
We found 27 publicly available scATAC-seq datasets and included them within scATAC.Explorer. 14 of 27 datasets contain cells from neuronal samples, 8 from immune cell samples, 6 from hematopoietic, and an assortment of others such as kidney, lung, epithelial, and other tissue and organ samples. 15 of 27 datasets contain mice samples, 8 from human samples, 3 from human and mouse cell lines, and a single dataset each relating to common fruit fly and chicken samples. 8 of 27 datasets relate to diseases such as diabetes, acute myeloid leukemia, astrocytoma, glioma, basal cell carcinoma, and other cancers. The included datasets vary in size between 227 to 246,942 sequenced cells and 20,103 to 279,559 genomic regions per cell.
This study has created a comprehensive scATAC-seq database that can be queried by researchers interested in analyzing scATAC-seq datasets relating to a range of organisms, cell types, and diseases. scATAC.Explorer can easily be integrated into existing R and Python analysis pipelines, allowing researchers to quickly analyze scATAC-seq data from multiple sources. The diversity in available datasets can also be used to quickly test and validate new scATAC-seq analysis techniques and pipelines. As new datasets are published, they will be included through the use of a dataset submission process.

PrgmNr 1044 - A statistical framework to assess replicability of signals from trans-ethnic genome-wide association meta-analysis: Applications to smoking/drinking addiction traits using 3.4 million individuals

View session detail

Author Block: **C. Wang**, GSCAN - the GWAS and Sequencing Consortium of Alcohol and Nicotine Use; Penn State Univ., Hershey, PA
Disclosure Block: **C. Wang:** None.

Consortium studies often use genome-wide association meta-analysis (GWAMA) aggregate summary statistics from multiple studies to empower genetic discovery. It is a standard practice to replicate the association signals using an independent dataset. Yet, as discovery studies continue to grow larger and more diverse, it becomes difficult to identify a large enough replication sample, and more so for studies of non-European ancestry. Without replication, the identified association signals are much more likely to be spurious and confound downstream studies. To address this challenge, we propose a novel statistical framework RATES (Replicability Assessment in Trans-Ethnic Studies) to assess replicability without a replication sample. RATES first models genetic effect variations across studies using meta-regression with principal components of genome-wide allele frequencies as covariates and adjusts genetic effect heterogeneities due to ancestry. Next, RATES leverages the strength and consistency of residual association signals across variants and studies to calculate a â□□posterior probability of replicabilityâ□□, based on the rationale that replicable association signals tend to be significantly associated across multiple studies. A parametric bootstrap method was also developed to evaluate the p-values for PPR. We performed extensive simulations where 1) the genetic effects are homogeneous across ancestries, 2) the genetic effects are ancestry-specific, and 3) false-positive signals occur in some studies in the meta-analysis. We compared RATES with popular meta-analysis methods including the fixed effect (FE), random effects (RE and RE2) and binary effect (BE) meta-analysis, and meta-regression (MR-MEGA). We showed when outliers are present, only RATES yields correct type I error, while other methods (e.g., FE or RE2) can have > 4 folds inflated type I error. RATES also gives higher or comparable power in all scenarios, even for simulations that favor alternative methods. For variants with ancestry-specific effects, the power of RATES is 7% to over 400% higher compared to the 2nd best performing meta-analysis method. We further applied RATES to smoking/drinking addiction traits using 3.4 million individuals of different ethnic groups. As the first step, RATES confirmed that all sentinel variants reported have PPR>99%. When comparing the mean chi-square as converted from p-values, RATES yields chi-square values that are 9 % higher than the 2nd best method (RE2). Applying RATES to rare and low-frequency variants that are typically filtered out, we further identified novel signals of biological relevance in addition to GWAMA of common variants.

PrgmNr 1045 - Pleiotropic decomposition regression to characterize multi-trait genetic architecture

View session detail

Author Block: **J. Ballard**, L. O'Connor; Broad Inst. of MIT and Harvard, Cambridge, MA
Disclosure Block: **J. Ballard:** None.
Pleiotropy is pervasive and points to shared genetic architecture and associations as well as shared biology among many common diseases and complex traits. Approaches that begin to quantify pleiotropy include cross-trait LDSC to estimate genetic correlation (Bulik-Sullivan et al. 2015 Nat Genet), as well as an approach based on joint fine-mapping (Pickrell et al. 2016 Nat Genet). Applications of these methods include multi-trait meta-analysis for increased power (Turley et al. 2018 Nat Genet), causal inference, and clustering disease variants by their effect on related phenotypes (Udler et al. 2018 PLOS Medicine). However, these various methods exploit only partial knowledge of the effect size distribution. We developed pleiotropic decomposition regression (PDR), which partitions the heritability of multiple traits into components implying different types of pleiotropic relationships (e.g. correlated, uncorrelated, trait-specific). We fit our model using Fourier regression (Oâ□□Connor in press Nat Genet). In simulations, our method distinguished genetic architectures with different types of pleiotropy. We applied our model to pairs of 23 traits from the UK Biobank to quantify pleiotropy from three sources: correlation, shared heritability enrichments and LD enrichments (Gazal et al. 2017 Nat Genet), and uncorrelated pleiotropy. Many trait pairs had substantial pleiotropy even in the absence of genetic correlation; about half was explained by shared enrichments and half was not, pointing instead to shared pathways and cell types. We exploited these sources of pleiotropy to produce Bayesian effect size estimates that were better than the original GWAS data. By incorporating data from three additional traits, accuracy improved from r^2 = 0.097 to r^2 = 0.280 for T2D and r^2 = 0.088 to r^2 = 0.311 for asthma in held-out data. These improvements exceeded the performance of MTAG (r^2 = 0.177 and r^2 = 0.171) and the performance of bivariate analyses.

PrgmNr 1046 - DeepNull: Modeling non-linear covariate effects improves phenotype prediction and association power

View session detail

Author Block: **F. Hormozdiari**[1], Z. R. McCaw[2], T. Colthurst[1], T. Yun[1], N. Furlotte[2], A. Carroll[2], B. Alipanahi[3], C. Y. McLean[1]; [1]Google Hlth., Cambridge, MA, [2]Google Hlth., Palo Alto, CA, [3]Google, Inc., Palo Alto, CA

Disclosure Block: **F. Hormozdiari:** Major Stockholder/Ownership Interest; All authors have Google LLC stock as part of their compensation.. Salary/Employment; All authors are employees of Google LLC..

Genome-wide association studies (GWAS) are among the foundations of statistical genetics research, having detected thousands of variants associated with complex traits and diseases. A typical GWAS examines the association between genotypes and a phenotype of interest while adjusting for a set of covariates. Although recent work has recognized the importance of including covariate-interaction effects (e.g. age and sex interactions), no systematic approach has been described for detecting the appropriate non-linear functions to adjust for in GWAS. Identifying the appropriate transformations and interactions among a set of covariates is difficult due to the exponential number of possible interactions that can arise from a finite set of covariates and the infinite number of possible transformations of any given continuous covariate. In addition, the optimal number of covariate interactions is unknown a priori and requires evaluating many different possible covariate interactions.

We introduce DeepNull, a method that models non-linear covariate effects on phenotypes using a deep neural network and then includes the networkâ□□s prediction as a single additional covariate in GWAS. DeepNull can be applied in conjunction with any existing GWAS method. We investigate the potential of DeepNull for improving power in simulated and real data sets. First, using simulated data, we show that DeepNull increases statistical power by up to 20% while maintaining tight control of the type I error in the presence of interactions and non-linear covariate effects. Second, DeepNull obtains similar results to standard GWAS when the covariates have only linear effects on the phenotype. Third, DeepNull detects more genome-wide significant hits (i.e. independent lead variants) and loci (independent regions after merging hits within 250 Kbp together) than standard GWAS (i.e. BOLT-LMM) for ten phenotypes obtained from UK Biobank (n=370K). For glaucoma referral probability (GRP; Alipanahi et al. 2021 AJHG), DeepNull detects 46.1% more significant loci (38 vs. 26). For ApoB, DeepNull detects 8.5% more significant loci (217 vs. 200). The average improvement for significant loci across all ten phenotypes was 7.9%. Finally, we observed that for GRP, LDL, Calcium, and ApoB, DeepNull improves phenotype prediction accuracy (R2) by 83.4% (P-200), 40.3% (Pâ□□189), 23.9% (Pâ□□44), and 21.6% (Pâ□□75), respectively, suggesting DeepNull as an avenue for improving polygenic risk prediction. Overall, DeepNull improves phenotype prediction by 23.7% on average for the ten phenotypes analyzed. The preprint is available at *doi.org/10.1101/2021.05.26.445783*.

PrgmNr 1047 - SR-TWAS: Leveraging multiple reference panels to improve TWAS power by ensemble machine learning

View session detail

Author Block: **R. Parrish**[1], A. S. Buchman[2], P. L. De Jager[3], D. A. Bennett[2], M. P. Epstein[4], J. Yang[1]; [1]Dept. of Human Genetics, Emory Univ, Atlanta, GA, [2]Rush Alzheimerâ□□s Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL, [3]Columbia Univ Med Ctr, New York, NY, [4]Emory Univ, Atlanta, GA
Disclosure Block: **R. Parrish:** None.

A transcriptome-wide association study (TWAS) is a popular technique for integrating reference transcriptomic data with GWAS data to conduct gene-based association studies. Existing methods often assume a single reference panel of transcriptomic and genetic data from a relevant tissue for model training and determination of expression weights to be integrated with GWAS data for TWAS. However, multiple reference panels of a given tissue often exist; the Religious Orders Study (ROS), Rush Memory and Aging Project (MAP), and Genotype-Tissue Expression project (GTEx) all profile transcriptomic data of prefrontal cortex tissue. We anticipate that leveraging multiple reference panels can improve the performance of existing TWAS tools such as TIGAR and PrediXcan compared to the standard use of a single reference panel.

Here, we develop a novel TWAS method to leverage multiple reference panels using an Ensemble Machine Learning technique of Stacked Regression to form optimal linear combinations of GReX prediction models trained from multiple reference panels of the same tissue type. We refer to this method as Stacked Regression based TWAS (SR-TWAS), which is expected to improve GReX prediction accuracy and TWAS power over existing methods by leveraging multiple reference panels with increased effective training sample sizes.

We validated our SR-TWAS method by simulation studies using real whole genome sequencing (WGS) genotype data from GTEx V8 and ROS/MAP, as well as real application studies leveraging reference panels of the GTEx V8 and ROS/MAP with transcriptomic data of brain frontal cortex tissue. Under simulation scenarios with various proportions of true causal eQTL SNPs per gene, p_{causal}=(0.001, 0.01, 0.05, 0.1), and simulated gene expression heritability, h_e^2=0.2, SR-TWAS had up to 29% higher average gene expression prediction accuracy and up to 76% higher TWAS power than TIGAR using a single reference cohort (either GTEx V8 or ROS/MAP).

In real application studies, SR-TWAS using both GTEx V8 and ROS reference panels outperformed TIGAR using either reference panel alone with the highest prediction accuracy in the MAP cohort. Increased power was demonstrated in an application TWAS of Alzheimer's disease using public GWAS summary statistics from the International Genomics of Alzheimer's Project, where SR-TWAS identified 11 independent risk genes compared to 2 identified by TIGAR using only the ROS/MAP reference panel.

Our SR-TWAS tool will be publicly available on Github, providing a useful resource for leveraging multiple public transcriptomic databases to increase GReX prediction accuracy and TWAS power for mapping risk genes of complex diseases.

PrgmNr 1048 - Genetic and environmental correlations between complex phenotypes differ by race/ethnicity and sex

View session detail

Author Block: **M. Elgart**[1], A. W. Manichaikul[2], X. Guo[3], N. Franceschini[4], B. M. Psaty[5], C. Isasi[6], S. S. Rich[2], J. I. Rotter[7], R. Kaplan[8], R. D. Hernandez[9], S. Redline[10], Trans-Omics for Precision Medicine (TOPMed) Consortium, T. Sofer[11]; [1]Brigham and Women's Hosp., Harvard Med. Sch., Boston, MA, [2]Univ Virginia, Charlottesville, VA, [3]Harbor-UCLA, Torrance, CA, [4]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [5]Univ. of Washington, Seattle, WA, [6]Albert Einstein Coll. of Med., New York, NY, [7]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [8]Albert Einstein Coll. of Med., Bronx, NY, [9]San Francisco, CA, [10]Brigham and Women's Hosp., Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [11]Harvard Med. Sch., Boston, MA

Disclosure Block: **M. Elgart:** None.

Rationale

Phenotypic correlations have genetic and environmental components. Previous studies used data from large biobanks to estimate genetic correlations between various phenotypes yet were limited to single race/ethnicity and sex combined. The computational approaches used to estimate genetic correlations are either computationally challenging (e.g. genetic restricted maximum likelihood, which uses individual level-data) and inapplicable to large cohorts, or fast (e.g. linkage disequilibrium score regression), but prone to inaccuracies when there is genetic heterogeneity between the target sample, reference LD panel, and the population used to compute genome-wide summary statistics.

Methods

We derived a closed form, computationally efficient, solution to estimate genetic and environmental correlation coefficients within the Haseman-Elston framework using individual level-data. We systematically interrogated heritabilities and genetic correlations between all phenotype-pairs of 28 blood pressure, lipids, blood counts, inflammation, and anthropometrics phenotypes from individual-level TOPMed dataset totaling 33,959 individuals, stratified by race/ethnicity. In Hispanic/Latino participants from HCHS/SOL we estimated genetic and environmental correlations between 61 phenotypes from 11 domains (diabetes, cardiovascular, blood pressure, kidney, lipids, lung, sleep, anthropometrics, iron, RBC, and WBC), further stratified by sex. Finally, we performed domain-level enrichment analysis.

Results

We estimated over 2000 genetic and environmental correlations between phenotype-pairs in the multi-ethnic TOPMed and the Hispanic/Latino datasets. Stratifying by race/ethnicity, 2% (Blacks, ~8K participants), 22% (Hispanics/Latino, ~8K participants) and 16% (Whites, ~16K participants) of the total estimated genetic correlations and heritabilities were significant only in that group. In HCHS/SOL, the genetic and environmental components of the phenotypic correlations differed between phenotype domains. For example, the association between diabetes and anthropometric, sleep, and lung domains were driven by genetic correlations, while the association between diabetes and cardiovascular and blood pressure domains were driven by shared environment. Domain correlation enrichment also differed by sex. E.g., sleep and anthropometric domains were enriched in genetic correlations only in males, while diabetes and lipids genetic correlation enrichment was only observed in females. While additional independent validation is needed, these results can be useful in prioritizing precision medicine initiatives

PrgmNr 1049 - Enrichment for gene-by-environment candidate SNPs with variance locus analysis: Extension of vQTL to binary traits in the Pan-UK Biobank

View session detail

Author Block: **X. Tong**, A. A. Motsinger-reif; NIH, Durham, NC
Disclosure Block: **X. Tong:** None.
Despite increasing awareness that gene-by-environment (GxE) interactions explain a portion of the missing heritability in complex diseases, statistical power limitations and computational tractability have limited the identification of GxE effects on disease outcomes. Variance quantitative trait locus (vQTL) analysis based on double linear modeling (DLM) or double generalized linear modeling (DGLM) has recently been used to overcome genome and exposome dimensionality issues. vQTL analysis allows the identification of GxE loci without assessing environmental exposures, based on the rationale that GxE unaccounted for in a genome-wide association study (GWAS) can result in heterogeneous residual variance across allele dosages in quantitative phenotypes.
In this study, we propose variance loci analysis (VLA) based on DLM augmented by squared SNP dosage. Through simulations, we demonstrated that VLA exhibits increased power to detect GxE candidates when direct environmental effects on the phenotype are weak. Furthermore, we show that DGLM is overly generic with squared dosage and overly restrictive without it. More importantly, we demonstrate that by properly adjusting covariates, VLA expands the vQTL approach to non-quantitative traits such as binary disease outcomes. VLA also has computational tractability superior to DLM/DGLM-based vQTL.
This extension to non-quantitative traits enables the reevaluation of publicly available genome-wide data to identify candidate variants likely involved in GxE interactions. We performed VLA on ~7,000 health- and disease-related phenotypes derived from the UK Biobank and cataloged the results analogous to the GWAS catalog. We show that our method enriches for variants across functional categories (e.g., eQTL and intergenic variants) and known environmental interactions without the need for environmental exposure data.
We used our most significant variance loci in type 2 diabetes (T2D), cardiovascular disease (CVD), and body mass index (BMI) to perform GxE analysis in the Personalized Environmental and Genes Study (PEGS), a North Carolina-based cohort (n=19,672) with extensive exposome (n=9,414) and whole genome sequencing (n=4,737). Our results confirm that the top VLA-selected candidates for CVD, T2D and BMI from the UK Biobank enriched the detection of genome-exposome-wide significant GxE variants in PEGS and demonstrate that the approach can be applied for both quantitative and binary outcomes.

PrgmNr 1058 - Long-Term systemic expression and cross-correction ability of HMI-203: Investigational gene therapy candidate for mucopolysaccharidosis type II or Hunter Syndrome

View session detail

Author Block: **K. Patel**[1], L. Smith[1], J. Gingras[2], A. Tzianabos[1], L. A. Schuman[1], V. Zhivich[1], M. Kivaa[1], D. Faulkner[1], A. Sengooba[1], L. Behmorias[1], T. Seabrook[1], S. Dollive[1], N. Avila[1], J. Zhao[1], Y. White[1], J. Newman[1], S. Woodcock[1], S. Smith[1], O. Francone[1], A. Seymour[2]; [1]Homology Med.s, Inc., Bedford, MA, [2]Homology Med.s, Bedford, MA

Disclosure Block: **K. Patel:** Salary/Employment; Homology Medicines Inc..

Mucopolysaccharidosis type II (MPS II), or Hunter syndrome, is a rare X-linked lysosomal storage disorder caused by mutations in the iduronate-2-sulfatase (*IDS)* gene, resulting in loss of I2S activity leading to systemic (peripheral organs and central nervous system (CNS)) toxic lysosomal accumulation of glycosaminoglycans (GAGs). GAGs are large polysaccharides made of repeating disaccharide units responsible for providing structure and hydration to the cell. The disease results in skeletal dysplasia, joint stiffness, organomegaly, airway obstruction and, in severe cases, neurocognitive deficits. Hunter syndrome occurs in approximately 1 in 100,000 to 1 in 170,000 males, and causes significantly reduced lifespan, with the severe form leading to life expectancy of 10 to 20 years The proposed therapeutic mechanism of gene therapy candidate HMIâ□□203 is based on both intracellular expression and synthesis of active I2S, as well as high levels of expression and secretion of active I2S enzyme to support cross correction. Herein, we report preclinical data where a single intravenous dose of HMI-203 delivering human *IDS* via a rAAVHSC vector in the MPS II murine model resulted in dose-dependent and long-term transduction, *IDS* expression and I2S enzymatic activity in the evaluated tissues, e.g., liver, brain and serum through 52 weeks post-dose. A significant correlation was observed between liver and serum I2S activity, suggesting that the liver was likely the major contributor to the elevated levels of active I2S in the serum. The circulating I2S protein in the serum was functionally active (i.e., 90 kDa form) and cross-correction activity via a mannose-6-phosphate receptor dependent pathway was demonstrated using an *in vitro* competition assay. The robust and broad *IDS* tissue expression, along with demonstrated cross-correction significantly reduced GAG heparan sulfate (GAG-HS) to wild type (WT) levels in all evaluated organs associated with the disease, cerebrospinal fluid (CSF) and urine. In addition, lysosomal-associated membrane protein-1 (LAMP1) levels were significantly reduced to WT-like levels in the peripheral organs and CNS tissues. Of note, positive and significant correlations were observed between reduction in GAG-HS and LAMP1 levels in the CNS and brain and CSF GAG-HS levels, suggesting that CSF GAG-HS levels could be indicative of overall brain GAG and lysosomal burden levels in the clinic. Taken together, we have demonstrated that HMI-203 combines transduction and expression with the potential for cross-correction. These HMI-203 IND-enabling studies support HMI-203 as a gene therapy candidate for the treatment of MPS II.

PrgmNr 1059 - Tasimelteon Safely and Effectively Improves Sleep in Smith Magenis Syndrome: results from a Double-Blind Randomized Trial Followed by an Open-Label Extension

View session detail

Author Block: **C. Polymeropoulos**[1], J. Brooks[1], S. P. Smieszek[1], E. Czeisler[1], M. Fisher[1], M. Gibson[1], K. Kite[1], C. Xiao[1], S. H. Elsea[2], G. Birznieks[1], M. H. Polymeropoulos[1]; [1]Vanda Pharmaceuticals Inc., Washington, DC, [2]Baylor Coll. Med., Houston, TX
Disclosure Block: **C. Polymeropoulos:** Salary/Employment; Vanda Pharmaceuticals Inc..
Smith-Magenis Syndrome (SMS; OMIM #182290) is a rare genetic disorder that results from an interstitial deletion of 17p11.2 and, in rare cases, from a retinoic acid induced 1 (*RAI1*) gene variant (Slager et al 2003). Currently, the prevailing theory is that there is an underlying circadian pathophysiology causing sleep disturbances in these patients, as they exhibit low overall melatonin concentrations and abnormal timing of peak plasma melatonin concentrations. This abnormal inverted circadian rhythm is estimated to occur in 95% of individuals with SMS (Boone et al., 2011; Spruyt et al., 2016). To assess the efficacy of tasimelteon, a melatonin receptor agonist, to improve sleep in SMS, a 9-week, double-blind, randomized, two-period crossover study was conducted at four U.S. clinical centers. Genetically-confirmed SMS patients, aged 3 to 39, with sleep complaints participated in the study. Patients were assigned to treatment with tasimelteon or placebo in a 4-week crossover study with a one week washout between treatments. Eligible patients participated in an open label study and were followed for > 3 months. Improvement of sleep quality (DDSQ50) and total sleep time (DDTST50) on the worst 50% of nights were primary endpoints. Secondary measures included actigraphy and behavioral parameters. Over three years, fifty-two patients were screened and twenty-five patients completed the randomized portion of the study. DDSQ50 significantly improved over placebo (0.4, p=0.0139) and DDTST50 also improved (18.5 min, p=0.0556). Average sleep quality (0.3, p=0.0155) and actigraphy-based total sleep time (21.1 min, p=0.0134) improved significantly, consistent with the primary outcomes. Patients treated for â¥ 90 days in the open label study showed persistent efficacy. Adverse events were similar between placebo and tasimelteon. Tasimelteon safely and effectively improved sleep in SMS. The 17p11.2 deletion encompasses *RAI1*, leading to haploinsufficiency, which is considered the primary cause for most features of SMS, including dysregulation of the molecular clock via its effect on *CLOCK* expression. ChIP-Chip and reporter studies suggest that RAI1 binds, directly or in a complex, to the 1st intron of *CLOCK*, enhancing transcriptional activity, resulting in reduced *CLOCK* expression in SMS patient-derived cells (Williams et al 2012). The results of this study suggest that treatment with a the circadian regulator can, in part, ameliorate the circadian deficiencies caused by RAI1 haploinsufficiency, providing further evidence of a critical role for RAI1 in the regulation of circadian rhythms.

PrgmNr 1060 - Unravelling African genomes: Whole-genome sequencing of 1000 Nigerian samples spanning 50 tribal groups provides new insights into diversity and admixture

View session detail

Author Block: **C. O'Dushlaine**[1], B. Ballew[1], C. Palmer[1], 54gene Research Team, NCD-GHS Consortium, B. Salako[2], O. Nash[3], S. Fatumo[4], O. Salako[5], O. Osifo[1], Z. Moddibo[1], A. Ado-Wanka[1], D. Attipoe[1], G. Eze-Echesi[1], A. Yakubu[1], J. Popoola[1], A. Ene-Obong[1]; [1]54gene, Washington, DC, [2]Nigerian Inst. of Med. Res., Lagos, Nigeria, [3]Ctr. for Genomics Res. and Innovation, Abuja, Nigeria, [4]H3Africa Bioinformatics Network (H3ABioNet) Node, Ctr. for Genomics Res. and Innovation, NABDA/FMST, Abuja, Nigeria, [5]Coll. of Med. Univ. of Lagos, Lagos, Nigeria
Disclosure Block: **C. O'Dushlaine:** Salary/Employment; 54gene.
The lack of adequate representation of diverse genomes in human genomics research may limit insights that can be made about variants influencing disease susceptibility and trait variability across populations. We are helping to address this gap by performing germline whole genome sequencing of a Nigerian cohort. Nigeria represents one of the most diverse and populous regions on earth, with a population of over 200 million and over 250 unique tribal groups. We coordinate data generation in Lagos with analysis by staff around the world by leveraging cloud resources and deploying a scalable, robust, portable pipeline for alignment and variant calling. We present results from an initial round of whole-genome sequencing of ~1000 subjects from 50 tribal groups in Nigeria. We describe patterns of variation across tribes including variants of different functional classes and frequencies. We survey patterns of autozygosity across groups and compare these to 1000 Genomes samples. We highlight genetic distances between tribes and reveal evidence of admixture with European and northern African populations. We compare frequencies within our dataset to those reported in publicly available data (e.g. 1000 Genomes) for specific loci of clinical utility, e.g. those associated with drug response, highlighting noteworthy differences. Lastly, we find widespread, tribe-specific differences in allele frequency for medically-relevant variation, underscoring the importance of variant discovery and replication in non-European ancestry cohorts. Our results add to the growing body of genomic data from diverse populations, investigating understudied groups and the unique opportunities for discovery that they represent. We highlight opportunities for precision medicine, and reveal insights about variants of most clinical importance within and between human populations.

PrgmNr 1061 - *NOTCH3* p.Arg1231Cys is Present in 1 in 92 Pakistani and Associated with Stroke

View session detail

Author Block: **J. L. Rodriguez-Flores**[1], S. Khalid[2], J. L. Marchini[1], A. Rasheed[3], D. Liu[1], K. Mahmood[4], J. Backman[1], I. Qureshi[5], B. Ye[1], B. Kalesan[1], I. Haider[4], N. Ahmed[6], W. Iqbal[7], S. Anjum[7], M. Hussain[7], A. Kamal[8], C. O'Dushlaine[1], C. Van Hout[1], M. Cantor[1], G. Tzoneva[1], J. Overton[1], J. G. Reid[1], A. Baras[1], N. Verweij[1], L. A. Lotta[1], G. Coppola[1], K. Karalis[1], A. Economides[1], Regeneron Genetics Center, A. R. Shuldiner[1], D. Salaheen[2,3]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Columbia Univ., New York, NY, [3]Ctr. for Non-Communicable Diseases, Karachi, Pakistan, [4]Dow Univ. of Hlth.Sci. and Civil Hosp., Karachi, Pakistan, [5]King Edward Med. Univ. and Mayo Hosp., Lahore, Pakistan, [6]Dept. of Neurology, Liaquat Natl. Hosp., Karachi, Pakistan, [7]Lahore Gen. Hosp., Lahore, Pakistan, [8]Dept. of Neurology, Aga Khan Univ. Hosp., Karachi, Pakistan

Disclosure Block: **J.L. Rodriguez-Flores:** Salary/Employment; Regeneron Genetics Center LLC.

Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL) is an autosomal dominant Mendelian disorder characterized by early onset of migraine with aura, recurrent stroke, and dementia. Pathogenic CADASIL variants either add or remove a cysteine (Cys+/-) residue in one of 34 epidermal growth factor like repeats (EGFR) in the extra-cellular domain (ECD) or *NOTCH3*. Exome-wide association analysis of 4,882 stroke cases and 6,094 controls recruited in the Pakistan Genomic Resource (PGR) from Pakistan identified one such variant, p.Arg1231Cys, associated with subcortical stroke; p value 2.18e-8, odds ratio (OR) 2.97, 95% confidence interval (CI) 2.03 to 4.35, minor allele frequency (MAF) 7.1e-3. Analyses of the larger PGR cohort comprising of 80,000 participants identified additional heterozygous and homozygous carriers of this variant; call back studies of the carriers and their family members identified a high mortality in family members and a high prevalence of stroke. The Cys allele was found to disrupt a highly conserved domain (91% overall sequence identity between human and mouse), was predicted deleterious by PolyPhen2 (score 0.843 of 1), and was risk-increasing (cases MAF 0.016, controls MAF 0.0053). The p.Arg1231Cys variant was observed at a similar MAF in other South Asian populations sequenced by Regeneron Genetics Center, and present but orders of magnitude rarer in European populations. Despite rare prevalence in Europe, p.Arg1231Cys was associated with ischemic stroke in 450 thousand UK Biobank (UKB) participants; p value 8.8e-4, OR 3.38, CI 1.65 to 6.94, MAF 2.0e-4. In addition, p.Arg1231Cys was associated with multiple brain MRI phenotypes relevant to CADASIL in a 40K subset of UKB, such as mean diffusivity in the external capsule; p value 5.41e-10, OR 1.4, CI 0.96 to 1.8, MAF 2.8e-4. Consistent with CADASIL pathogenicity, a burden test limited to Cys+/- variants in the NOTCH3 ECD (including p.Arg1231Cys) strengthened associations in both Pakistan (subcortical stroke p value 1.5e-10, OR 3.39, CI 2.32 to 4.91) and UKB (ischemic stroke p value 9.3e-8, OR 3.38, CI 1.74 to 2.98). In both cohorts, p.Arg1231Cys was the most common Cys+/- variant in the NOTCH3 ECD. Taken together, these findings have major implications for precision medicine in South Asia, given that an estimated 1 in 92 (over 20 million of 1.9 billion) individuals are carriers for this variant and are at approximately 3-fold elevated risk for stroke. Our estimates suggest that around 2% of strokes in Pakistan may be attributable to NOTCH3 p.Arg1231Cys.

PrgmNr 1062 - A high-resolution panel for uncovering repeat expansions that cause ataxias

View session detail

Author Block: **Y-C. Tsai**[1], Z. McEachin[2,3], M. van Blitterswijk[4], J. Ziegle[1]; [1]Pacific BioSci.s, Menlo Park, CA, [2]Dept. of Cell Biology, Emory Univ., Atlanta, GA, [3]Lab. for Translational Cell Biology, Emory Univ., Atlanta, GA, [4]Dept. of NeuroSci., Mayo Clinic, Jacksonville, FL
Disclosure Block: **Y. Tsai:** Salary/Employment; Pacific Biosciences.
The hereditary ataxias are a group of rare neurological diseases with similar symptoms. Many of these ataxic syndromes are caused by expansions of short tandem repeat (STR) in a number of different genes. Molecular genetic testing to accurately determine the genetic cause of known ataxias is often employed to support clinical diagnoses. Advances in therapeutic strategies (e.g., antisense oligonucleotides) to target repeat expansions underscore the importance of understanding the genetic context and sequence complexity of ataxic repeat expansions. Further highlighting the importance of molecular genetic testing, several studies have shown that repeat sequence interruptions in certain ataxia expansions play important roles in modifying the penetrance of the disease and age of onset. PCR and Southern blotting assays are currently the most employed methods in commercially available ataxia repeat expansion panels for clinical testing. Although these electrophoresis-based methods could detect repeat expansions above pathogenic threshold, accurate sizing of the repeat expansion is difficult to achieve when the length of repeat sequence is longer than a few hundred bases. Sequence interruption information is also not available with these approaches. We have recently developed an ataxia expansion panel using the PacBio No-Amp targeted sequencing approach to capture and sequence repeat expansion loci associated with fifteen ataxia diseases. The method utilizes CRISPR-Cas9 nuclease and pairs of guide RNAs to excise DNA fragments containing the repeat sequences within ataxia genes. This approach eliminates PCR amplification artifacts, amplification bias, and preserves native DNA for base modification detection. In this study, we sequenced samples with known or unknown diagnosis for ataxia with the No-Amp targeted sequencing panel utilizing PacBio highly accurate long reads - HiFi reads. The high accuracy of HiFi reads provides both certainty in sizing of the repeat expansion and repeat sequence interruption within the expansion sequences. Sequencing results demonstrate the potential of using this repeat expansion panel for eventual genetic testing. As additional ataxia, and related neurological diseases, caused by STR expansions are discovered and studied, the No-Amp targeted sequencing panel could be expanded to include additional targets. The ability to multiplex samples from different patients also makes the method a potentially cost-effective option for molecular genetic screening in the future.

PrgmNr 1063 - Deployment of clinical whole genome sequencing in support of more than 1,000 resource-limited patients: four years of the iHope Program

View session detail

Author Block: **E. Thorpe**[1], J. Ortega[1], K. Robinson[1], M. Jones[2], D. Basel[3], T. Williams[1], C. Brown[4], K. Vaux[5], A. Lumaka Zola[6], F. Sirchia[7], D. Masser-Frye[8], R. Foster-Bonds[9], A. Taylor[10], H. Reid[11], W. R. Wilcox[12], M. DueÃ±as Roque[13], M. D. Lah[14], M. Cornejo-Olivas[15], H. Pena Salguero[16], L. Davis Keppen[17], A. Serize[18], P. Magoulas[19], A. Vanderver[20], D. Henry[21], M. Shinawi[22], A. Malhotra[23], J. Avecilla[1], A. Warren[1], D. L. Perry[24], J. Belmont[25], R. J. Taft[26]; [1]Illumina, Inc., San Diego, CA, [2]Hosp. Infantil de las Californias, Tijuana, Mexico, [3]Children's Hosp. of Wisconsin, Milwaukee, WI, [4]Le Bonheur Children's Hosp., Memphis, TN, [5]Brain Vitals, San Diego, CA, [6]Ctr. For Human Genetics Univ. of Kinshasa UNIKIN Faculte de Medecine, Kinshasa, Congo, Democratic Republic of the, [7]Citta della Salute e della Scienza di Torino Genetica Medica, Turin, Italy, [8]San Diego Regional Ctr., San Diego, CA, [9]Rare Genomics Inst., Los Angeles, CA, [10]Al Jalila Childrens Hosp., Dubai, United Arab Emirates, [11]Canterbury Hlth.Lab., Christchurch, New Zealand, [12]Emory Univ, Atlanta, GA, [13]Hosp. Natl. Edgardo Rebagliati Martins, Lima, Peru, [14]IUSM, Indianapolis, IN, [15]Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [16]Padrino Fndn., Todos Los Santos, Mexico, [17]Sanford Hlth., Sioux Falls, SD, [18]South Miami Hosp., Miami, FL, [19]Baylor Coll. of Med. and Texas Childrenâ□□s Hosp., Houston, TX, [20]The Children's Hosp. Of Philadelphia, Philadelphia, PA, [21]UCSF, San Francisco, CA, [22]Washington Univ. Sch. of Med., St Louis, MO, [23]Illumina, Inc, San Diego, CA, [24]San Diego, CA, [25]Baylor Coll. of Med., Houston, TX, [26]Illumina Inc, San Diego, CA
Disclosure Block: **E. Thorpe:** Major Stockholder/Ownership Interest; Illumina. Salary/Employment; Illumina.

Patients with a suspected genetic disease are often unable to obtain a timely molecular diagnosis, and those in resource-limited locations face even greater challenges. Clinical whole genome sequencing (cWGS) shows promise as a comprehensive test which may shorten the diagnostic odyssey regardless of setting. The iHope Program is a philanthropic effort to provide cWGS to patients who are unable to obtain precision testing due to resource-limitations.

From June 2016 through June 9, 2021, 1004 individuals pursued cWGS test through the iHope Program. Cases were received from 23 partner iHope clinical sites spanning seven countries. Forty percent of cases (n=403) were received from global partners in Mexico (n=205), Peru (n=93), Italy (n=50), Democratic Republic of Congo (n=40), New Zealand (n=10), and the United Arab Emirates (n=5). Most testing was performed on duos and trios. Proband phenotypes were complex, with nervous system, head and neck, skeletal, eye, and digestive the five most frequently identified Human Phenotype Ontology root ancestor terms.

Variants were reported in 67% (n=677) of cases, of which 40.5% (n=407) conferred a definitive molecular diagnosis. Reported variants per case ranged from 0 to 5, and in 33 cases (3.3%), multiple molecular diagnoses were observed. Variants spanned 468 unique single genes. Of 1020 reported variants, a majority were nuclear SNVs or MNVs (n=693, 67.9%), followed by CNVs (n=175,17.2%), small indels (n=127, 12.5%), short tandem repeats (n=12, 1.2%), mitochondrial SNVs (n=10, 1%) and uniparental disomy (n=2, 0.1%). Copy number variants ranged in size from 3 kb to full aneuploidies. In fifteen individuals from eleven families, findings were suggestive of a structural chromosomal rearrangement.

At least ninety days after cWGS report delivery, a clinical utility survey was requested of the ordering clinician to assess effects on care and management. To date, surveys have been obtained for 581 patients (58%), representing one of the largest cWGS clinical utility datasets in a pediatric outpatient population. Data collection is ongoing, but initial analysis indicates that cWGS results prompted follow-up such as imaging, laboratory or physiological testing, referral for specialty consultation or

other evaluations in 40% (233/581) of patients. In 56.6% (329/581), cWGS results contributed to counseling about prognosis, recurrence risks, reproductive screening/testing options and screening/testing recommendations or options for family members. These findings suggest that deployment of cWGS in support of resource-limited patients is tractable globally and can have a substantial impact on patient management.

PrgmNr 1066 - Accurate multi-cancer detection and typing by genome-wide cell-free DNA profile data mining

View session detail

Author Block: **H. Che**[1], T. Jatsenko[1], L. Lenaerts[1], L. Dehaspe[2], L. Vancoillie[2], B. Nathalie[2], K. Van Den Bogaert[1], I. Parijs[2], D. Lambrechts[1], A. Coosemans[1], S. Tejpar[1], K. Punie[2], S. Hatse[1], P. Neven[3], H. Wildiers[3], E. Wauters[3], C. Dooms[3], A. Wozniak[1], P. Schoffski[1], D. Timmerman[1], P. Vandenberghe[1], F. Amant[1], J. R. Vermeesch[4,1]; [1]KU Leuven, Leuven, Belgium, [2]UZ Leuven, Leuven, Belgium, [3]Univ. Hosp. Leuven, Leuven, Belgium, [4]Univ. Hosp. of Leuven, Leuven, Belgium
Disclosure Block: **H. Che:** None.
Cell-free DNA (cfDNA) consists of a mixture of DNA fragments released from various dying cells of the organism, providing unique access to molecular information about the pathological process and tissue of origin. cfDNA analysis in cancer patients holds a great promise for non-invasive disease management. Existing analysis paradigms concentrate on detecting genomic alterations but have rarely been designed to integrate biological properties of cfDNA. Therefore, they lacked the generalized ability to explore a range of diseases. Here we developed a generic approach, coined GIPSeqcluster for unbiased identification of genome-wide signatures in shallow whole-genome sequencing (0.1x) cfDNA data. This approach enables detection of different physiological and pathological conditions and opens alternative avenues for cfDNA investigation. With enhanced detection of informative molecular signals in cfDNA, we uncovered cancer-specific patterns through unsupervised clustering that stimulated accurate cancer signal detection and cancer type mapping via supervised learning. We applied GIPSeqcluster on cfDNA samples from healthy controls (n=367) and patients with different types of hematological cancers (n=238) and solid tumors (n=517) of varying stages. Surveying the large volume of cfDNA profiles, we revealed malignancy type-specific clusters that empowered disease localization. Moreover, we identified intriguing common cfDNA patterns in patients with cancer. Classification yielded an overall accuracy of 80.11% in discriminating malignant from healthy samples, resulting in sensitivities ranging from 13.04% to 94.97% at 95% specificity across different cancer types and stages. Increased identification of cancer signals was obtained with the AUC of 0.988 and 0.804, compared to AUC of 0.926 and 0.717 using copy number alterations only analysis, from patients with hematological and solid cancer, respectively. Performing cancer type classification in samples with cancer-associated signal, the disease type could be mapped in 70% of cases. Our approach offers a widely applicable strategy for non-invasive cancer detection and highlights crucial properties of cfDNA in improving our understanding of tumor biology.

PrgmNr 1067 - A novel method identifies liquid biopsy markers based on relative chromatin accessibility

View session detail

Author Block: **M. Mehrmohamadi**, P. Taklifi, F. Palizban; Univ. of Tehran, Tehran, Iran, Islamic Republic of
Disclosure Block: **M. Mehrmohamadi:** None.
Liquid biopsy testing refers to minimally invasive cancer detection using markers in bodily fluids. Dying cells, including those originating from tumors, shed their DNA into the blood and contribute to a pool of circulating fragments called cell-free DNA (cfDNA). Tumor-originating fragments harbor genetic and epigenetic markers of cancer. However, the majority of cfDNA fragments originate from blood cells during normal turnover, leading to a high background of non-marker fragments that necessitate deep next-generation sequencing for cancer detection. Thus, targeted sequencing panels are designed to enrich and sequence only small portions of the genome harboring cancer markers. Most of the currently existing targeted sequencing panels cover recurrent genetic aberrations. To improve the utility and sensitivity of liquid biopsies, focus has shifted to epigenetic markers at tissue-specific regulatory regions. Here, we propose a novel strategy to select marker regions based on data of assay for transposase accessible chromatin followed by sequencing (ATAC-seq). Since accessible chromatin regions are tissue-specific and undergo higher degradation, they are manifested as local decrease in sequencing depth and fragment length in cfDNA. Our pipeline integrates ATAC-seq data from TCGA tumors and normal blood samples to measure a relative chromatin accessibility metric. In each cancer, regions with a high fold-change in relative accessibility between tumor and blood (log ATAC-seq normalized count in cancer/blood > 1.5) are selected since they are expected to show fragmentation differences in cfDNA of healthy individuals compared with cancer patients. We then use the selected regions as input features into our machine-learning pipeline and build a classifier distinguishing each cancer from other cancers in the TCGA ATAC-seq dataset. We applied this pipeline to 6 common cancers from the TCGA and obtained high specificity and sensitivity in held-out data (average AUC=0.99). Regions in promoters of many known tissue-specific genes were among our identified markers such as *KLK2* in prostate, *NOXO1* in colorectal, *TBC1D9* and *ERBB4* in breast, *CLD18* and *STK31* in gastric, *NR1H4N* in kidney, and *SFTPB* and *SFA3* in lung cancers. These observations confirm the biological relevance of our findings. We then directly validated our marker regions in publicly available deep whole-genome sequencing cfDNA data from cancer patients**.** Overall, we propose a novel marker selection strategy that can be used either alone or in integration with other genetic and epigenetic markers for efficient liquid biopsy panel design.

PrgmNr 1068 - Distinct intratumoral oncogenic mechanisms between single cells in primary Ductal Carcinoma In Situ exposed by Primary Template-directed Amplification

View session detail

Author Block: **J. Zawistowski**[1], R. A. Carter[1], J. R. Marks[2], L. Simpson[2], D. M. Arvapalli[1], T. V. Morozova[1], E. S. Hwang[2], J. A. West[1]; [1]BioSkryb, Inc., Durham, NC, [2]Duke Univ. Med. Ctr., Durham, NC
Disclosure Block: **J. Zawistowski:** Major Stockholder/Ownership Interest; BioSkryb, Inc.. Salary/Employment; BioSkryb, Inc..
Ductal Carcinoma In Situ (DCIS) is a neoplastic proliferation of the ductal epithelium that can transition to invasive ductal carcinoma (IDC) and ultimately lead to metastatic dissemination of tumor cells. To design therapeutic strategies targeting this exploitable DCIS-to-IDC junction in the progression of the disease it is of paramount importance to define the specific genomic lesions that are driving the transition. The complete complement of genomic changes within rare intratumoral clones that have gained capacity to invade and contribute to the progression of disease are not detectable by bulk sequencing. We have therefore employed Primary Template-Directed Amplification (PTA), a whole-genome amplification (WGA) technology, to ascertain the genome sequence of single cells dissociated from a surgical resection of a patient tumor comprised of both DCIS and IDC features. The core technology of PTA attenuates the size of daughter amplicons to limit subsequent amplification, thereby yielding high breadth of coverage and uniformity with significantly diminished allelic skewing to accurately call single nucleotide variation (SNV) and copy number variation (CNV). As initial genomic insight into the DCIS to IDC transition in this patient we performed PTA followed by whole genome sequencing on 31 single cells that were immunoenriched for EpCAM by FACS, including 5 single cells cells that were derived from an ipsilateral breast stromal control biopsy. As validation of our technology and experimental design we identified prototypical copy number alterations, including chromosomal loss at regions harboring *TP53*, *BRCA2*, and *RB1* tumor suppressors. We observed striking heterogeneity in single-cell CNV profiles, whereby multiple bins of EpCAM-enriched cells emerged: a bin with no apparent copy number alterations, those with different combinations of prototypical CNVs, as well as a class of cells harboring copy number losses at novel, non-prototypical loci. While cataloging SNV, we identified a common driver mutation in DCIS/IDC, *PIK3CA* H1047R. Intriguingly, the single cells harboring *PIK3CA* H1047R lacked copy number alterations. This suggests distinct mechanisms of oncogenesis: some cells within the tumor proliferate uncontrollably due to loss of key tumor suppressor regulation, while in other single cells a missense mutation in a key signal transduction node affecting downstream MAPK-mediated cell proliferation and AKT-mediated survival signaling is sufficient to drive unchecked growth. We are currently extending these findings to comprehensively define prototypical and novel SNV in these individual cells, in both genic and intergenic space.

PrgmNr 1069 - Full-length transcriptome analysis of liver cancer using a long-read sequencing technology

View session detail

Author Block: **H. Kiyose**; The Univ. of Tokyo, Hongo, Bunkyo Ku, Japan
Disclosure Block: **H. Kiyose:** None.
Background: Various transcripts can be generated from a single gene by alternative splicing. These transcripts lead to variants of proteins with diverse functions that can play an important role in carcinogenesis. Therefore, the analysis of full-length transcripts should be significant for cancer research. However, the majority of previous transcriptome studies have used microarrays or short-read sequencing technologies, and due to the difficulty of predicting full-length transcripts with these technologies, abnormalities at the transcript level have been overlooked. One promising approach is the direct observation of full-length transcripts using long read sequencing technology. However, due to the high error rate, the analysis method has not been established and its application to cancer research has been limited. **Results:** In this study, we developed an analysis pipeline named SPLICE to analyze full-length cDNA sequencing data. SPLICE removed possible errors around splicing junction sites and pseudo fusion transcripts by considering alignment and mapping errors. Validation with MCF-7 cell line and PCR showed the reliability of our analysis. Using this method, we analyzed cDNA sequences from 42 pairs of hepatocellular carcinoma (HCC) and matched non-cancerous liver with MinION (Oxford Nanopore technologies). Our analysis detected 46,663 transcripts from the protein-coding genes in the HCCs and the matched non-cancerous livers, of which 5,366 (11.5 %) were novel. We identified 767 and 531 novel exons from the novel transcripts in HCCs and matched-livers, respectively. Interestingly, about 40 % of the novel exons were derived from transposable elements (TEs), which was a significantly higher proportion than the TE-derived known exons. Comparison of expression levels identified 9,933 differentially expressed transcripts (DETs) in 4,744 genes. Importantly, 746 genes with DET were not found by the gene-level analysis. We also detected cancer-related genes with TE-derived exons. In the analysis of transcripts from hepatitis B virus (HBV), HBx-human TE fusions were found to be overexpressed in the HCCs. Furthermore, we found 6 novel recurrent fusion genes. By comparing HBV- with HCV-related livers, significant differentiations in immune-related genes were detected. **Conclusion:** These results suggest that long-reads sequencing technologies allow us to analyze full-length transcripts, and show the importance of splicing variants in carcinogenesis.

PrgmNr 1070 - Functional genomics uncover novel regulatory genomic regions associated with gene expression and survival in multiple myeloma

View session detail

Author Block: **L. Bui**, H. Natri, B. Argente, A. J. Gutierrez, M-I. Chung, L. Peter, J. Keats, N. E. Banovich; Translational Genomics Res. Inst., Phoenix, AZ
Disclosure Block: **L. Bui:** None.
Multiple Myeloma (MM), a malignancy of plasma cells, is the second most common hematological cancer and is associated with a poor prognosis. MM displays a sex disparity in incidence with men having a 1.5-fold higher risk in developing MM than women. Recent GWAS and meta-analyses have identified 176 risk loci associated with MM susceptibility; of these, many are located within or adjacent to the regulatory regions, indicating a role in transcriptional regulation. However, the regulatory mechanisms underlying how these risk loci contribute to tumor etiology, progression and outcome are poorly understood.
In this study, we investigate the role of genetic variation (somatic or germline) on gene expression changes in MM patients. First, we analyzed whole-genome sequencing (WGS) from peripheral blood and WGS and RNA-seq from baseline tumor specimens of 693 participants from the CoMMpass study, a longitudinal clinical trial of the Multiple Myeloma Research Foundation. We performed joint and sex-stratified analyses and detected 7,737 variants associated with changes in gene expression, i.e. expression quantitative trait loci (eQTLs). Among the identified eQTLs, 33.75% exhibit sex-specific effects and 1,034 variants are associated with survival in the CoMMpass cohort. Second, we selected 159 eQTLs, corresponding to 110 unique target genes (eGenes), and validated their functions in transcriptional activity in MM cell lines (2 females, 2 males) using CRISPR interference screens (CRISPRi) with a scRNA-seq readout. These variants include 60 male specific eQTLs, 27 female specific eQTLs and 43 eQTLs associated with survival from the CoMMpass cohort. We detected reduction in expression levels of 68 out of the 110 eGenes in the cells containing the gRNA targeting eQTLs, confirming the roles of these eQTLs in gene expression regulation. We analyzed global transcriptional changes and identified potential downstream targets that may play essential roles in tumor etiology and progression in MM. Among eQTL/eGene pairs with global transcriptional effects are PSAP (Prosaposin), encodes a protein that interacts with CD74 and may play a role in MM carcinogenesis and CCT8 (a subunit of Chaperonin-containing T-complex protein 1) that is known to promote the migration and invasion of squamous carcinoma. We identify differences of the sex specific variants on gene expression regulation in different sexes. In conclusion, we have performed computational analysis and functional validation to identify new regulatory regions associated with gene expression changes in MM, thus providing novel targets for the development of personalized medicine approaches for better treatments.

PrgmNr 1071 - Genetic map of human brain folding and links to developmental pathways and disease

View session detail

Author Block: S. J. Loomis[1], **B. Sun**[1,2], F. Pizzagalli[3], A. Zhu[3], D. Dixon[3], T. Islam[3], I. ba Gari[3], N. Shatokhina[3], H. Runz[1], N. Jahanshad[3], P. Thompson[3], C. D. Whelan[1]; [1]Biogen, Cambridge, MA, [2]BHF Cardiovascular Epidemiology Unit, Dept. of Publ. Hlth.and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, [3]Imaging Genetics Ctr., Mark and Mary Stevens Neuroimaging and Informatics Inst., Keck Sch. of Med. of USC, Univ. of Southern California, Marina del Rey, CA
Disclosure Block: **B. Sun:** Salary/Employment; Biogen.
The human brain is a complex organ underlying many cognitive and physiological processes as well as a wide range of diseases. Genetic associations with brain structure are emerging but studies looking at specific local brain folding remain under-explored. Here we carried out the first detailed large-scale genome-wide association studies (GWAS) of regional brain cortical folding (sulci and gyri) in UK Biobank. We discovered 272 brain folding region associations across 77 regions at *p*-8 in GWAS (discovery n=26,530) which replicated at *p*

PrgmNr 1074 - Trans-ancestry imputation and exome sequencing of more than 1 million individuals identifies genetic variation protecting against SARS-CoV-2 infection and predicts individuals at risk for severe COVID-19 outcomes

View session detail

Author Block: **J. Kosmicki**[1], J. Horowitz[2], A. Damask[1], J. Mbatchou[1], A. Marcketta[1], L. Dobbyn[1], D. Sun[1], J. D. Backman[1], N. Banerjee[1], A. Verma[3], A. Baras[1], E. Stahl[1], J. L. Marchini[1], G. R. Abecasis[1], M. A. R. Ferreira[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Regeneron Pharmaceuticals, Tarrytown, NY, [3]Dept. of Genetics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA
Disclosure Block: **J. Kosmicki:** Salary/Employment; Regeneron.
COVID-19 symptoms vary widely, ranging from asymptomatic in some patients to fatal in others. Elucidating the host genetics of COVID-19 holds the potential for understanding both susceptibility to SARS-CoV-2 infection as well as heterogeneity in patient presentation and outcome. Prior work focused on identifying common variants associated with COVID-19 susceptibility and severity, but little has been done to explore the entire allele frequency spectrum of genetic variation, from common to rare exonic variants. Here, we present the largest trans-ancestry exome sequencing study of COVID-19 to date in 586,713 individuals, with a larger set of 1,012,636 individuals with imputed data across 7 studies and 5 continental ancestries.

Through exome sequencing of 21,820 COVID-19 cases and 564,893 controls, we did not identify any rare variants after Bonferroni correction (PDISP3 (*P*=2e-8; OR=1.8Â±0.3), *MARK1* (*P*=3e-9; OR=38.4Â±16.9), and *TLR7* (*P*=4e-8; OR=4.5Â±2.2). Despite having a 100x larger sample size, we could not replicate a previous reported role for rare variants in the interferon pathway (*P*=0.59).

Our larger GWAS of 56,841 cases and 955,795 controls found 11 loci (PACE2, the primary cell receptor for the SARS-CoV-2 spike protein (*P*=4.5e-13; OR=0.6Â±0.08; EUR MAF=0.003). Using RNA-seq, rs190509934 reduced ACE2 expression by 39% (P=3e-8), supporting the hypothesis that reduced ACE2 expression protects against SARS-CoV-2 infection.

Lastly, we developed a polygenic risk score (PRS) to predict hospitalization and severity of COVID-19. Among those of European ancestry, individuals with the top 10% PRSs are 1.8-fold more likely to be hospitalized (*P*=6e-11) and 1.58-fold more likely to be placed on a ventilator or die from COVID-19 (*P*=7e-10). These associations hold in other non-European populations (albeit with decreased power) and after accounting for known clinical risk factors.

Our data represents the most comprehensive survey of common and rare exonic variation associated with COVID-19 identifying new loci and polygenic risk scores that predict severity of COVID-19.

PrgmNr 1075 - Rare variant analyses in 239,395 whole exome and whole genome sequenced participants of the UK Biobank reveals novel genetic associations with renal function and chronic kidney disease

View session detail

Author Block: **S. Li**[1], B. Mautz[2], P. JongHanne[1], S. Guo[3], H. Hejase[4], G. Abdel-Azim[5], N. Azhar[1], V. Krishna[1], B. Sarver[6], T. R. Parrado[7], G. Rajagopal[8], H. Makimura[1], M. Breyer[1], D. F. Reilly[9], M. Black[10]; [1]Janssen, Spring house, PA, [2]Nashville, TN, [3]North Wales, PA, [4]Janssen Res. & Dev., Willow Grove, PA, [5]Ellicott City, MD, [6]Fort Washington, PA, [7]Spring House, PA, [8]Janssen R&D, Spring House, PA, [9]Janssen Pharmaceuticals, Boston, MA, [10]Janssen Res. & Dev. (Johnson & Johnson), Spring House, PA
Disclosure Block: **S. Li:** Salary/Employment; Janssen.
Genome-wide association studies have identified common genetic variants associated with chronic kidney disease (CKD), but the burden of rare loss-of-function (LoF) or pathogenic/likely pathogenic (P/LP) variants has not been well characterized. We performed gene-/region-based and variant association analyses for 5 renal function biomarkers (eGFR estimated from serum creatinine and/or cystatin-C, BUN, UACR) and 5 CKD endpoints (ESRD and stage4/5 CKD, CKD defined by biomarkers and/or diagnoses from NHS data, Cystic) in 239,395 UKB participants of genetically-assessed European ancestry and with whole exome (WES, n=171,172) or whole genome sequencing (WGS, n=121,019). For each trait, we fit a genome-wide regression model and tested for association using REGENIE V2.0, adjusting for age, sex, 10 principal components of ancestry, assessment center and BMI, where appropriate. For gene-based analyses, we generated 15 models to collapse ClinVar-classified P/LP, VEP(LOFTEE)-predicted putative LoF and deleterious variants predicted by 16 *in silico* scores (SIFT, Polyphen, BayesDel, etc.) from dbNSFP 4.1c. The WGS data further enabled annotation of promoter/enhancer variants, which were incorporated into collapsing models for gene-based association. In participants with WES, we identified 30 and 11 genes associated with â¥2 biomarkers and â¥1 CKD endpoint across collapsing models (FDRPKD1/2, COL4A3/4, CUBN, IFT140 were associated with both biomarkers and CKD. Association analyses also highlighted other genes including: *COL4A1, CST3, LAMC1, LRP2, SLC22A2, SLC34A3, SH2B3*. Variant-level analyses further informed impact on protein, e.g. the *SLC22A2* association signal was mainly driven by a frameshift (rs8177505) with lowering effects on eGFR (p=1.2e-27, beta=-6.2, MAF=0.12%). Exome-wide variant analyses revealed 25 genes (eg. *PDILT-UMOD*) with variant associations (p3 biomarkers or â¥1 endpoint, including 2 that were also implicated from the gene-based analyses (*COL4A4* and *CUBN*). Analyses of WGS allowed for sequence level validation of exome derived findings and the identification of additional variants not captured in WES. This study provides a framework for the assessment of the genetic landscape of kidney disease. The results validated known genes and identified potential novel associations with renal function.

PrgmNr 1076 - Novel genetic associations for rare diseases with GWAS and trans-ethnic analysis of self-reported medical data

View session detail

Author Block: **S. S. Shringarpure**[1], W. Wang[1], Y. Jiang[1], A. Acevedo[2], D. Dhamija[1], B. Cameron[1], A. Jubb[1], P. Yue[3], 23andMe Research Team, L. sarov-Blat[2], R. Gentleman[1], A. Auton[1]; [1]23andMe, Sunnyvale, CA, [2]GlaxoSmithKline, Collegeville, PA, [3]23andme, Sunnyvale, CA
Disclosure Block: **S.S. Shringarpure:** Salary/Employment; 23andMe Inc..
Nearly 7000 rare diseases are known, and though each disease affects a few people, the total population prevalence of rare diseases is estimated to be 3.5-5.9%. A key challenge in the study of rare disease genetics is assembling large case cohorts for well-powered studies. Here we demonstrate use of large-scale self-reported rare disease data, combined with genetic data collected through the 23andMe direct-to-consumer platform, to study 33 rare diseases and identify genetic associations through GWAS. We developed web-based questionnaires, and gathered self-reported data on rare diseases from a cohort of over 1.6 million genotyped research-consented individuals. To reduce mis-reporting and maximize coverage, we used an autocomplete mechanism including 7000 rare diseases. We validated the approach through simulations and replication of known rare disease associations. In simulations based on genotypes from 4,957,230 European individuals, we show that GWAS can recover genome-wide significant associations in monogenic rare diseases for a variety of architectures. In rare diseases with known genetic associations, we reidentified 29 associations at a genome-wide significance level (p-value

PrgmNr 1077 - Common and rare variant analysis of 21K psoriasis cases and 623K controls identifies novel, protective associations in several genes in the type 1 interferon pathway

View session detail

Author Block: **J. Horowitz**[1], L. Gurski[1], D. Sun[1], L. Khrimian[2], M. LeBlanc[1], M. Jones[1], N. Banerjee[1], A. Verma[3], Regeneron Genetics Center, Geisinger DiscovEHR Collaboration, D. J. Rader[4], M. D. Ritchie[3], J. T. Elder[5], A. Franke[6], M. A. Ferreira[1], K. A. Siminovitch[1], A. Baras[1]; [1]Regeneron Genetics Ctr., Tarrytown, NY, [2]Regeneron, Tarrytown, NY, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]Univ of Pennsylvania, Philadelphia, PA, [5]Univ MICHIGAN, Ann Arbor, MI, [6]Kiel Univ., Kiel, Germany
Disclosure Block: **J. Horowitz:** Salary/Employment; Regeneron Genetics Center.
Psoriasis is a complex autoimmune disease resulting in chronic inflammation and hyperproliferation of the skin. The aberrant immune response associated with psoriasis is mediated by pathogenic T cells, which are activated, in part, by type 1 interferons (IFNs). Prior large-scale analyses of psoriasis cases focusing on common genetic variants have implicated >63 loci, including genes in the IFN signaling pathway. However, large-scale analysis of rare exonic variation is lacking.

To study the contribution of both common and rare variants to psoriasis risk, we performed whole-exome sequencing and meta-analysis of 20,810 psoriasis cases and 623,159 controls of EUR and AFR ancestry across 6 cohorts. Common variant analysis replicated 44 significant and independent associations in known psoriasis loci, including *IL23R*, *TYK2*, *IL12B*, *HLA-C*, and *DDX58*, among others. Rare-variant gene-burden analysis of putative loss-of-function (pLoF) and/or predicted-deleterious missense variants (IFIH1 (OR=0.74 [0.68, 0.81], p=4.1E-12), which encodes a pathogen sensor that activates IFN production, and *TRIM65* (OR=0.63 [0.50, 0.79], p=4.8E-5), which encodes a ubiquitin ligase that binds and activates IFIH1. We find the protective *TRIM65* association is driven by a rare, predicted-deleterious missense variant (rs202175254, AAF=0.1%) in the IFIH1-TRIM65 binding domain. Further, we find a nominally significant, protective association for the burden of rare pLoFs in *DDX58* (OR=0.76 [0.49, 0.89], p=6.7E-3), which encodes a second pathogen sensor that activates IFNs. This *DDX58* protective pLoF association helps confirm direction of effect at this known psoriasis locus.

Consistent with inhibition of IFNs being protective in psoriasis, we also found a significant and novel gene-burden association between increased odds of psoriasis and pLoFs in *ADAR* (OR=2.29 [1.68, 3.12], p=1.4E-7), which encodes a protein that suppresses IFNs and in which partial LoFs have been associated with Aicardi-GoutiÃ¨res syndrome, an inherited disorder that features over-production of IFNs.

Collectively, these results represent the largest rare-variant exome-sequencing analysis of psoriasis, to date. Future experiments will characterize effects of these pLoFs on protein expression and/or function, and further analysis will determine whether an IFN gene signature can identify a clinically-relevant subset of psoriasis patients who would therapeutically benefit from IFN inhibition.

PrgmNr 1078 - Investigating genetic and phenotypic associations for 168 blood metabolites in 120K UK Biobank participants

View session detail

Author Block: **A. Nag**[1], Q. Wang[2], A. Harper[1], D. Paul[1], A. Reznichenko[3], B. Challis[3], K. Carss[1], S. Petrovski[1]; [1]Ctr. for Genomics Res., AstraZeneca, Cambridge, United Kingdom, [2]Ctr. for Genomics Res., AstraZeneca, Waltham, MA, [3]Translational Sci. and Experimental Med., AstraZeneca, Gothenburg, Sweden

Disclosure Block: **A. Nag:** Salary/Employment; AstraZeneca.

In this study, we accessed the large-scale metabolomics, exome sequencing and phenomics data from the UK Biobank (UKB) to investigate gene-metabolite and metabolite-phenotype relationships. Blood metabolites (N=168) were profiled by Nightingale Health in ~120,000 UKB participants, >90% of whom had exome sequences and all had data on ~16,000 clinical traits.

We explored genetic associations with blood metabolites by two complementary approaches: (i) single-variant analysis, and (ii) gene-level collapsing analysis, using a linear regression model, adjusted for age, sex and BMI. For the single-variant analysis, we tested ~3.2 million variants under dominant and recessive models. For the gene-level collapsing analysis, the aggregate effect of variants in each gene was tested using 11 different models, including ones that focused on rare (MAFOur analyses provide a rich catalogue of significant (p-8) associations: 10,461 variant-metabolite, 970 gene-metabolite, and 127,947 metabolite-phenotype relationships. This includes well-established, biologically plausible associations such as variants in *PAH* with phenylalanine levels [beta=1.2; p-300] and the concentration of intermediate-density lipoprotein particles with type 2 diabetes [beta=-1.5; p-300]. These data may also provide insights into underlying biological mechanisms: for instance, the observed metabolite signature for mutations in a gene that is a known drug target (e.g, *HSD17B13*) can indicate the metabolic profile expected with desirable therapeutic response.

The catalogue of genetic and phenotypic relationships for blood metabolites, which will expand further once metabolomics data becomes available in the entire UKB cohort of ~500,000 subjects, represents an excellent resource to better understand mechanisms underlying complex human diseases.

PrgmNr 1079 - Practical implementation of polygenic risk scores and absolute risk score estimation across diverse ancestry groups

View session detail

Author Block: **R. Moore**, M. Golden, D. Thompson, F. Riveros-Mckay, P. Donnelly, G. McVean, V. Plagnol, M. E. Weale; Genomics plc, Oxford, United Kingdom
Disclosure Block: **R. Moore:** Salary/Employment; Genomics plc.
Polygenic risk scores (PRS) have generated considerable translational interest. Yet, most validation efforts focus on assessing relative rather than absolute risk scores (ARS), even though ARS are required for clinical decision making. ARS validation experiments are typically based on a single large cohort split into training/testing and rarely incorporate PRS. While such approaches typically generate calibrated ARS within the testing dataset, they do not properly capture the complex biases inherent to each healthcare context or account for environmental differences between countries and ethnicities. Consequently, the robustness of the ARS across different contexts is largely unknown.
To address these gaps, we derived a framework to combine ethnicity-specific disease baselines from a range of country-specific surveys, which capture social determinants of health, with ancestry-adjusted PRS (European OR per 1SD 1.87, 2.10, 1.51 and 2.09 respectively) for breast cancer, prostate cancer (PC), cardiovascular disease (CVD) and type 2 diabetes (T2D). We validated these ARS in independent datasets, computing calibration summary statistics, including the standard incidence ratio (SIR), calibration slope and intercept, and the integrated calibration index.
We find that inclusion of an ethnic specific baseline captures substantial ARS variability not captured by the PRS, particularly for PC, where an UK African and Caribbean baseline results in calibration (0.99-1.34 95% CI SIR) whilst the UK average baseline results in strong miscalibration (2.24-3.02 95% CI SIR). The extent of the calibration varied, with challenges arising for T2D and CVD, whose incidence has fluctuated across time and location in the US over the last decades. For T2D, baselines date from 1997-2019 but prospective testing data date from 1987-1999, resulting in miscalibration for White males (1.35-1.62 95% CI SIR). For CVD, baselines for myocardial infarction and fatal heart disease date from 2004-2011 and ischemic stroke from 1999, but prospective testing data date from 1986-2000, resulting in miscalibration for White females and males (0.66-0.92 and 1.04-1.31 95% CI SIR respectively).
We demonstrate that with appropriate data it is possible to translate genetic risk into clinically meaningful ARS that robustly replicate in diverse contexts. Our results also demonstrate the challenges arising from variation across ethnicity, geography and time and the need for population-relevant information on which risk prediction tools are to be applied.

PrgmNr 1103 - Characterization of molecular and motor phenotypes in CGG knock-in mice with CRISPR mediated deletion of the trinucleotide repeat

View session detail

Author Block: **C. M. Yrigollen**[1], L. Ohl[1], E. Lim[1], S. Zheng[2], K. Brida[3], Y. Chen[1], B. Simpson[1], B. L. Davidson[4]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Temple Univ., Philadelphia, PA, [3]Univ. of Alabama at Birmingham, Birmingham, AL, [4]The Children's Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **C.M. Yrigollen:** None.
Fragile X-associated Tremor/Ataxia Syndrome (FXTAS) is a late onset neurodegenerative disorder that is caused by a premutation allele (55-200 CGG repeats) in the 5â untranslated region of the *FMR1* gene. *Fmr1* CGG knock-in (KI) mice with approximately 130 CGG repeats exhibit motor and memory impairments and are used to model FXTAS. We evaluated a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Cas9 based therapy for its ability to correct *FMR1* trinucleotide repeats in the KI mouse *in vivo.*

Two guideRNAs (gRNAs) were used to target the CGG repeats for deletion upon Cas9 mediated cleavage. Dual AAV vectors containing Cas9 and the gRNAs were injected into neonatal mouse pups at P0 to P1 bilaterally into the cerebral ventricles. Motor performance was evaluated at 12-15 weeks, 28-30 weeks and 52-54 weeks of age by accelerating rotarod. Motor impairment occurs in KI mice by 28 weeks, and results in quicker falls from the apparatus than is observed in WT littermates. Mice that were treated with the CRISPR constructs performed better than untreated knock-in mice, $p = 0.0041$ and were not statistically different their WT littermates at 28-30 weeks of age. At the late stage rotarod the *Fmr1* treated mice continued to perform similarly to the WT animals, but statistical differences were no longer observed between treatment groups.

Robust expression of SpCas9 and GFP were observed in the cortex, hippocampus, striatum, and cerebellum, of treated knock-in animals by qRT-PCR, but no significant reduction in Fmr1 was observed. These results are consistent with only a subset of cells needing to be transduced and edited to achieve a phenotypic rescue. Long read sequencing was performed to characterize the on-target outcomes of this gene editing strategy. We identified both expected and unexpected editing events (i.e., deletion of CGG repeats and vector sequence integration, respectively), highlighting the importance of fine mapping editing outcomes of this nascent gene therapy technology.

Our study is the first to demonstrate *in vivo* editing of expanded CGG repeats in *Fmr1* using CRISPR. This results in a therapeutic benefit, rescuing motor deficits present in aged KI mice. Long read sequencing and qRT-PCR provide insights into the variation of edits being achieved and how often unexpected editing events occur at the target locus. These results are important indicators that CRISPR mediated gene-editing is worth further development for treatment of FXTAS and other Fragile X-associated disorders.

PrgmNr 1104 - Upregulation of disease compensatory gene via CRISPR activation in muscular dystrophy mice

View session detail

Author Block: **J. Cheng-Zhang**[1,2], M. Johnson[2], V. Alday[2,3], A. Munir[2], A. Arockiaraj[1,2], O. Akinyele[2], J. F. Alcorn[4], R. D. Nicholls[1,2], D. U. Kemaladewi[1,2]; [1]Dept. of Human Genetics, Graduate Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [2]Div. of Genetic and Genomic Med., Dept. of Pediatrics, Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA, [3]Dept. of NeuroSci., Kenneth P. Dietrich Sch. of Arts and Sci., Univ. of Pittsburgh, Pittsburgh, PA, [4]Div. of Pulmonology, Dept. of Pediatrics, Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA

Disclosure Block: **J. Cheng-Zhang:** None.

Merosin-deficient congenital muscular dystrophy type 1A (MDC1A) is caused by mutations in the *LAMA2* gene encoding laminin-Î±2, an extracellular protein that is essential for skeletal muscle and Schwann cell functions. MDC1A patients with complete loss of expression never achieve independent ambulation and often do not reach adolescence, whereas those with partial deficiency exhibit milder phenotypes. Furthermore, there is a heterogeneity in the mutations found in the MDC1A patient population, which hampers the translation of genome editing approaches. The objective of this project is to develop a mutation-independent approach to treat MDC1A by upregulating a compensatory gene called *LAMA1* using CRISPR activation. We packaged a catalytically inactive SaCas9, 2XVP64 transcriptional activators and guide RNAs targeting the *Lama1* promoter into two adeno-associated viruses 9 (AAV9). Subsequently, we injected the AAV9 into the temporal vein of neonatal dy2j/dy2j, a mouse model with slightly reduced Lama2 and moderate phenotypes, or dyW/dyW, which have very low Lama2 and limited lifespan. We assessed the neuromuscular functions and histopathology to evaluate the therapeutic efficacy. We showed robust LAMA1 upregulation in skeletal muscle and peripheral nerves, leading to improvement in the dystrophic histopathology and functions, suggesting the therapeutic potential of this strategy. Current work involves the generation of a â□□mini-and-mightyâ□□ system by incorporating miniaturized transcriptional activators and utilization of single-AAV9 delivery system. Subsequently, we will evaluate the efficacy and safety of this strategy in the severe dyW/dyW mice. Successful completion of this project will open up new therapeutic avenues based on upregulation of compensatory gene in many diseases.

PrgmNr 1105 - Neuronal gene interaction:*Scn8a*, *Kcna1* and *Kcnq2*

View session detail

Author Block: **S. Hill**[1], P. Jafar-Nejad[2], F. Rigo[2], M. Meisler[3]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Ionis Pharmaceuticals, Carlsbad, CA, [3]Univ Michigan, Ann Arbor, MI
Disclosure Block: **S. Hill:** None.
Gain-of-function mutations in the neuronal voltage-gated sodium channel gene *SCN8A* are a cause of developmental and epileptic encephalopathy (DEE). Modification of *Scn8a*-DEE in the mouse by interaction with a mutation of *Gabra2* was recently described (Yu et al, *Epilepsia* 61:2847-2856, 2020). The severity of *Scn8a*-DEE can be ameliorated by administration of an antisense oligonucleotide (ASO) that reduces expression of *Scn8a* (Lenk et al*, Ann. Neurol.* 87:339-346, 2020). This ASO also rescues survival of mice with mutation of the sodium channel *Scn1a.* The proteins encoded by *SCN1A* and *SCN8A* are concentrated at the axon initial segment of the neuron, where they contribute to the generation of action potentials. To evaluate genetic interaction between *Scn8a* and other channels located at the axon initial segment, we examined mice with mutations in the potassium channel genes *Kcnq2* and *Kcna1*. Mice with homozygous mutations of *Kcna1* and *Kcnq2* were treated on postnatal day 2 by intracerebroventricular injection of a dose of ASO that reduces the level of *Scn8a* transcript by 50%. Treatment with the *Scn8a* ASO resulted in significant delay in the age of seizure onset, as well as increased length of survival. A small effect was observed in mice with mutation of the synaptic protein LGI1. The data demonstrate that the severity of genetically determined seizure disorders can be modified by compensatory changes among co-localized ion channels. An ASO targeting *SCN8A* may therefore have therapeutic value in multiple types of genetic epilepsy.

PrgmNr 1106 - Adjunct treatment with glycogen synthase (GYS1) antisense oligonucleotides: Reduce glycogen in the Pompe disease mouse model

View session detail

Author Block: **L. Weiss**[1], M. Carrer[2], J. Lee[1], A. Shmara[1], J. Vu[1], C. Cheng[1], N. Raben[3], T. R. Grossman[2], J-n. Paymaan[2], V. E. Kimonis[1]; [1]Univ. of California, Irvine, Irvine, CA, [2]Ionis Pharmaceuticals, Carlsbad, CA, [3]NIH, Bethesda, MD

Disclosure Block: **L. Weiss:** None.

Pompe disease is a progressive myopathy resulting from the deficiency of acid a glucosidase. Enzyme replacement therapy (ERT) with recombinant human GAA works well in alleviating the cardiomyopathy; however, many patients continue to have progressive muscle weakness from muscle glycogen accumulation. Previous studies have provided proof of principle that knockdown of GYS1 mRNA by phosphorodiamidate morpholino oligonucleotide reduced glycogen. The lead from the screening over 150 ASOs were validated in a dose response study and the top 10 ASOs were screened in vivo. Three GYS1 ASOs showed the best tolerability and efficacy profile leading to knock down of GYS1 mRNA by approximately 50% of control. We performed an in vivo study of the toxicity and efficacy of the three GYS1 ASOs in Gaa -/- Pompe mice as monotherapy. The results showed over 16 doses (25mg/kg, once a week) of the best candidate, ASO#3 was safe and well tolerated. It significantly reduced muscle GYS1 mRNA levels, autophagy pathology, to a lesser extent glycogen content in muscle of Gaa -/- mice compared to the control groups, PBS and mismatch ASO. We are currently treating mice with a combination treatment of ASO#3 and ERT to maximize the therapeutic effect. After four months of treatment, the combination treatment group showed an upward trend of improvement in motor tests compared with ERT or ASO#3 alone groups.

PrgmNr 1107 - Transethnic evaluation of copy number burden in epilepsies from a genome-wide study of 34,132 subjects

View session detail

Author Block: L-M. Niestroj[1,2], **C. Leu**[1,3,4], D. Lal[1,5,4,2], Epi25 Consortium; [1]Genomic Med. Inst., Lerner Res. Inst., Cleveland Clinic, Cleveland, OH, [2]Cologne Ctr. for Genomics, Univ. of Cologne, Cologne, Germany, [3]Dept. of Clinical and Experimental Epilepsy, UCL Inst. of Neurology, London, United Kingdom, [4]Stanley Ctr. for Psychiatric Res., Broad Inst. of Harvard and M.I.T, Cambridge, MA, [5]Epilepsy Ctr., Neurological Inst., Cleveland Clinic, Cleveland, OH
Disclosure Block: **C. Leu:** None.
Rationale: Rare copy number variants (CNVs) are strongly implicated in the etiology of epilepsy. We previously published an association study of CNVs with different types of epilepsy in ~10k European epilepsy cases and ~6k ancestry-matched controls (Niestroj et al., 2019, *Brain*). We observed large differences in the CNV burden across epilepsy types and different CNV categories. We found that the established epilepsy-associated 15q13.3 deletion represents the strongest risk CNV for genetic generalized epilepsy (GGE), with a 34-fold risk for developing GGE. However, similar to most previous epilepsy genetics studies, we focused only on individuals with European ancestry. The distribution of CNVs that confer risk towards epilepsy and their effect sizes are not well established in non-European populations.
Methods: Using genome-wide genotyping data generated with the Illumina Global Screening Array-24 v1.0, we reanalyzed the current cohort of the Epi25 Consortium together with new genetic data, totaling 21,706 individuals with epilepsy and 12,426 ancestry-matched controls for the burden of rare CNVs.
Results: We identified 25,172 European- (EUR), 4,087 African- (AFR), and 1,824 Asian-ancestry (ASN) individuals in our study available for CNV burden analyses after quality control. Across all ancestries, we identified 305 CNV carriers of seven established epilepsy-associated CNV hotspots in 19,079 individuals with epilepsy and 70 carriers among 12,004 controls. Among all tested epilepsy subsyndromes and in line with previous evidence, we observed the highest CNV burden in individuals with GGE. The two most common recurrent CNVs, deletions at 15q11.2 and 16p13.11, known to be associated with GGE in European individuals, were also enriched in African and Asian individuals with GGE, with effect sizes similar to those in European individuals (e.g., 15q11.2: $P_{EUR}=6.89x10^{-5}$, $OR_{EUR}=2.44$ vs. $P_{AFR}=9.49x10^{-3}$, $OR_{AFR}=3.96$; or 16p13.11: $P_{EUR}=9.26x10^{-12}$, $OR_{EUR}=25.9$ vs. $P_{ASN}=6.58x10^{-3}$, $OR_{ASN}=26.4$). At the time of the conference, we will present additional analyses of different classes of CNVs and genome-wide CNV breakpoint association analyses that explore the similarities and possible differences between CNV architectures of European, African, and Asian ancestries.
Conclusions: Our study represents the largest and the first transethnic evaluation of the CNV burden in epilepsy. Our preliminary results indicate that established recurrent CNVs have similar effect sizes in different ethnicities. As CNVs are part of clinical genetic testing, our results will likely help alleviate health care disparities.

PrgmNr 1108 - Variant-to-gene-mapping of insomnia GWAS loci followed by a neuron-specific RNAi screen in *Drosophila melanogaster* implicates *PIGQ* at the â□□*WDR90*â□□ locus as a novel regulator of sleep

View session detail

Author Block: **S. F. Grant**[1,2], J. Palermo[3], S. Sonti[1], C. Lasconi[1], A. Chesi[1,2], D. R. Mazzotti[4], P. R. Gehrman[2], A. C. Keene[3]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Perelman Sch. of Medcine, Univ. of Pennsylvania, Philadelphia, PA, [3]Florida Atlantic Univ., Jupiter, FL, [4]Univ. of Kansas Med. Ctr., Kansas City, KS

Disclosure Block: **S.F. Grant:** None.

Insomnia is the most common sleep disorder and increases risk for subsequent psychiatric sequelae. In recent years, GWAS efforts have identified >200 insomnia loci, but the underlying causal mechanisms remain elusive. We elected to conduct the first variant to gene mapping for insomnia by identifying putative causal variants and their associated effector genes followed by assessing functional effects on sleep/wake regulation in the fruit fly, *Drosophila melanogaster*. Genome-wide significant signals identified from published insomnia GWAS (n>200) were included in this study. We integrated proxy SNPs in high linkage disequilibrium to each sentinel signal with both genome wide ATAC-seq and high-resolution promoter-focused Capture C data in order to implicate effector genes contacted directly by regulatory regions coinciding with these variants in iPSC-derived neural progenitor cells. The resulting putative effector gene list was subjected to a neuron-specific RNAi screen in *Drosophila melanogaster* followed by in-depth sleep phenotyping in order to assess the impact of such genetic perturbation on sleep/wake regulation. As a consequence, a number of short- and long-sleeping lines were derived. Most notably, multiple lines of evidence converged on the *PIGQ* gene at the â□□*WDR90*â□□ locus, which encodes phosphatidylinositol N-acetylglucosaminyltransferase subunit Q, an enzyme involved in the first step of glycosylphosphatidylinositol anchor biosynthesis. Knockdown of PIGQ significantly increased both daytime and nighttime sleep. Restricting knockdown to cholinergic neurons phenocopied pan-neuronal knockdown, supporting the notion that PIGQ modulates the function of excitatory neurons. To examine how PIGQ impacts sleep, we used video tracking prior to and following exposure to mechanical shaking of sleeping flies and measured the effect on sleep. No differences in arousal threshold were observed, suggesting that PIGQ regulates sleep duration, but not arousability which is considered an indicator of sleep depth. Together, these findings suggest PIGQ is selectively a regulator of sleep duration, but not sleep quality. Our ongoing studies seek to localize PIGQ function to specific regions within the brain, and to identify its impact on sleep intensity and function more precisely. These analyses strongly support a role for *PIGQ* at the â□□*WDR90*â□□ locus as a gene encoding a novel regulator of sleep, and thus a key contributor to the pathogenesis of insomnia. These results highlight the strength of leveraging multiple, converging approaches to characterize the genetic basis of insomnia based on the findings resulting from GWAS efforts.

PrgmNr 1111 - Somatic Mosaicism and Cardiovascular Disease: Uncovering Hidden Genetic Etiology with Clinical Implications

View session detail

Author Block: **V. Chander**[1], E. Venner[1], M. Murugan[1], M. Cohen[1], G. Metcalf[1], M-C. Gingras[1], C. Kovar[1], J. Hu[1], P. Panchal[1], V. Korchina[1], D. Muzny[1], D. Murdock[1], E. Boerwinkle[2], C. Ballantyne[3], R. Gibbs[1]; [1]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [2]UTHlth.Sch. of Publ. Hlth., Univ. of Texas, Houston, TX, [3]Dept. of Med./Div. of Cardiology, Baylor Coll. of Med., Houston, TX
Disclosure Block: **V. Chander:** None.
Cardiovascular disease (CVD) is the leading cause of death globally and constitutes a significant health burden. The identification of numerous genes associated with CVD has led to their incorporation into clinical genetic testing panels and yet the molecular basis of almost half of CVD cases remains unknown. A growing number of genetic disorders have been associated with somatic mosaicism, however, the prevalence and impact of somatic mutations in CVD have not been well-explored. Recent identification of the association of Clonal Hematopoiesis of Indeterminate Potential (CHIP), driven by somatic mosaicism, with increased risk for atherosclerosis suggests a role in CVD. Deep DNA sequencing can enable sensitive detection of somatic mutations and elucidation of their role in CVD.

We determined the prevalence of somatic mutations within a heart disease-enriched cohort (HeartCare), a genetic screen of >700 individuals, in ambulatory cardiovascular care clinics, targeting 158 genes associated with Mendelian CVD conditions. Using deep NGS to detect low-frequency disease-associated variants, we detected exonic somatic mutations in more than 10% of individuals, including a subset recurring in the same locus. To identify pathogenic somatic mutations, we filtered out common polymorphisms using population-level databases and narrowed down the deleterious variants to those supported by *in silico* pathogenicity prediction evidence. Among them, 82% had CVD indications but were negative for a germline mutation or LPA finding, thus without a clear genetic diagnosis.

Interestingly, we found links between somatic mutations in implicated genes (*ABCC6, ABCC9, SDHA, CACNA1C, CACNB2, COL5A1, COL5A2, MYH7, MYH11, RYR2*) and CVD events (e.g arrhythmia, high calcium score, heart attack, A-Fib) with instances where the expected gene-specific phenotype was observed, supporting causality. The implicated genes encode key proteins involved in drug and lipid transport, ATP-dependent transport in mitochondria, calcium channels, collagen synthesis, and actin-based motors. These genes have moderate to high expression in the cardiac tissue, including a subset expressed in hematopoietic stem cells with a potential role in CHIP. These results support the hypothesis that somatic mosaicism in CVD impairs pathways involved in intracellular transport and calcium channeling and may underly the link between clonal expansions and CVD conditions. Collectively, these results demonstrate the power of targeted NGS studies to reveal the impact of somatic mosaicism and suggest that somatic mosaicism represents an important etiology of CVD.

PrgmNr 1112 - Identifying the role of alternative splicing during human heart organogenesis

View session detail

Author Block: **K. Child**, T. Yankee, A. Wilderman, J. Cotney; UConn Hlth., Farmington, CT
Disclosure Block: **K. Child:** None.
Congenital Heart Defects (CHD) are one the most frequent congenital abnormalities affecting births worldwide. Many groups investigating CHDs have been focused on heart gene expression during fetal development and later. However, recent evidence has suggested that gene expression dynamics principally occurs during organogenesis, which happens before the fetal stage. This period encompasses the first eight weeks of human development and has been classified into 23 Carnegie Stages (CS). Our lab has published extensive work detailing the nature of gene expression during these early stages focusing on CS12 to 23 (4 to 8 post-conception weeks). Examination of gene co-expression across this window of development revealed RNA processing and splicing as a major module in this network. Furthermore, recent work has shown that alternative isoform utilization is the most dynamic in the heart and brain during fetal development which were the earliest stages surveyed. These results suggest alternative splicing and existence of novel isoforms during heart organogenesis could be important for normal heart development during the embryonic period. To identify alternative splicing events across development we reanalyzed short-read bulk gene expression from embryonic human heart. Consistent with results from fetal data, we found many differential splicing events between individual Carnegie stages within this period. These differentially spliced genes were enriched for functions related to heart development and abnormalities, however few overlapped with our previous differential expression results calculated at the composite gene level. Generation of *de novo* transcriptome assembly using this data revealed numerous previously unannotated isoforms, even amongst well studied heart related genes, many of which were differentially spliced during this developmental period. To confirm the existence of these novel isoforms we subsequently profiled the transcriptome of these same samples using Nanopore long-read sequencing. We validated a substantial number of these novel isoforms including a novel *TBX5* transcript which utilizes several previously unannotated 5â□□ exons. These exons could be important for post-transcriptional regulation of TBX5 and be an additional region to screen for in Holt-Oram Syndrome patients. Overall, our analysis identifies previously unknown transcript diversity in the developing human heart which could help explain previously undiagnosed genetic causes of CHD.

PrgmNr 1113 - Genetic deficiency of the AIM2 inflammasome mitigates *JAK2*p.V617F clonal hematopoiesis associated cardiovascular diseases risk

View session detail

Author Block: **Z. Yu**[1], T. Nakao[1], M. Uddin[1], A. Niroula[1], T. P. Fidler[2], S. M. Zekavat[3], C. J. Gibson[4], G. K. Griffin[5], A. R. Tall[2], P. Libby[5], S. Jaiswal[6], B. L. Ebert[4], A. Bick[7], P. Natarajan[8]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Columbia Univ. Irving Med. Ctr., New York, NY, [3]Yale Sch. of Med.; Broad Inst., Natick, MA, [4]Dana-Farber Cancer Inst., Boston, MA, [5]Brigham and Women's Hosp., Boston, MA, [6]Stanford Univ. Sch. of Med., Stanford, CA, [7]Vanderbilt Univ. Med. Ctr., Nashville, TN, [8]Massachusetts Gen. Hosp., Boston, MA

Disclosure Block: **Z. Yu:** None.

Clonal hematopoiesis of indeterminate potential (CHIP), a common condition among elderly where clonal expansion of blood cells harboring somatic mutations, has been associated with the risk of hematologic malignancy and cardiovascular diseases (CVD). *JAK2*^{V617F} is the most commonly observed point mutation in CHIP and confers the strongest CVD risk. Murine studies implicate NLRP3 inflammasome activation for *Tet2* CHIP, but both NLRP3 and AIM2 inflammasome activation for *JAK2*^{V617F} CHIP and atherosclerosis. Whether AIM2 inflammasome inactivation mitigates *JAK2*^{V617F} CHIP-associated CVD in human is unknown. We used whole exome sequencing data of blood DNA analyzed from 184,887 genetically unrelated participants in the UK Biobank without prevalent CVD and hematological cancers to identify those with *JAK2*^{V617F} mutations. We generated an *AIM2* gene expression reducing eQTL score using eQTL summary statistics from the eQTLGen consortium, where each score represents a genetic proxy for gene expression. We also extracted *IL6R* p.Asp358Ala genotype status, which was previously linked to reduced primarily *DNMT3A* and *TET2* CHIP-associated CVD risk. We examined the association between the *JAK2*^{V617F} status and incident CVD (myocardial infarction, coronary revascularization, stroke, or death) risk, and whether the *AIM2* eQTL score or *IL6R* p.Asp358Ala modifies that association. In the analyzed dataset, the mean (SD) age was 56.3 (8.0) years and 103,958 (56.2%) were women. We identified 114 (0.06%) individuals with *JAK2*^{V617F} mutation, 112 of whom carried large clones (allele fraction > 10%). During 10.8-year median follow-up, 18,771 incident CVD events were observed. *JAK2*^{V617F} was associated with 2.8-fold increased incident CVD risk (P $6.6*10^{-11}$) independent of age, sex, self-reported race, genotyping array, and first 10 principal components of ancestry. Decreased genetic predisposition to *AIM2* expression attenuated the risk among *JAK2*^{V617F} carriers (HR 0.27; 95% CI 0.09-0.84; P 0.023) but not those without the *JAK2*^{V617F} mutation (HR 1.00; 95% CI 0.98-1.01; P 0.73; $P_{interaction}$ 0.024). Restriction to large *JAK2*^{V617F} clones did not alter the results. Nor did carrier status of the *IL6R* p.Asp358Ala coding mutation exhibit interaction with *JAK2*^{V617F} status on CVD risk. Leveraging human genetic instruments, we observed that the *JAK2*^{V617F} mutation enhances CVD risk, but that genetically reduced *AIM2* expression attenuates this risk. Our human genetic findings corroborate recent murine observations and suggests that *AIM2* inhibition could specifically mitigate *JAK2*^{V617F} CHIP associated CVD. This hypothesis merits testing assessed in prospective clinical trials.

PrgmNr 1114 - *FHL5*, a Novel Coronary Artery Disease Gene, Regulates Smooth Muscle Cell Contractility

View session detail

Author Block: **D. Wong**[1], A. Turner[1], M. Khan[1], N. Perry[1], M. Palmore[1], M. Kuppusamy[1], M. Ottolini[1], L. Matic[2], U. Hedin[2], L. Ma[3], S. K. Sonkusare[1], M. Civelek[4], J. C. Kovacic[5,6], J. Bjorkegren[7,8], C. Miller[1]; [1]Univ. of Virginia, Charlottesville, VA, [2]Karolinska Inst., Stockholm, Sweden, [3]Columbia Univ, New York, NY, [4]Univ Virginia, Charlottesville, VA, [5]Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Victor Chang Cardiac Res. Inst., Darlinghurst, Australia, [7]Icahn Sch. of Med. at Mount Sinai, Charlottesville, NY, [8]Karolinska Inst.t, Huddinge, Sweden

Disclosure Block: **D. Wong:** None.

Coronary artery disease (CAD) is the leading cause of death worldwide. Genome-wide association studies (GWAS) have identified over 200 loci associated with CAD, many of which cluster into distinct functional pathways. For example, in the most recent meta-analysis of myocardial infarction (MI) and CAD, *UFL1-FHL5* (chr6q16.1) emerged as a novel CAD/MI locus. Despite additional genetic associations with multiple vascular pathologies, the underlying mechanism of this locus has not been elucidated. We first prioritized *FHL5* as the top candidate causal gene at *UFL1-FHL5* using two statistical fine-mapping approaches, Summary-level Mendelian Randomization and *coloc*. FHL5 expression is enriched in contractile mural cells (smooth muscle cells (SMCs) and pericytes) and potently downregulated in atherosclerotic arteries. Given its reported role as a cofactor, we hypothesized that FHL5 functions as a transcriptional regulator of SMC contractile pathways to impact vascular tone and disease risk. To test the functional effects of FHL5 overexpression in SMCs, we performed confocal-based calcium imaging and collagen gel contraction assays. FHL5 overexpression increased the basal calcium concentration 2.5x, which correlated with increased contractility. To interrogate the molecular mechanism of FHL5 function in SMCs, we mapped FHL5 binding sites using the Cleavage Under Targets and Release Using Nuclease (CUT&RUN) method. Genes harboring FHL5 binding sites were enriched for CAD relevant pathways, such extracellular matrix organization and TGF-beta signaling. Furthermore, FHL5 binding sites were also enriched for CAD/blood pressure (BP) risk variants, suggesting interactions between FHL5 and other CAD/BP loci. Finally, by performing weighted gene co-expression and key driver analyses in coronary arteries, we demonstrate that FHL5 possess regulatory potential *in vivo* and resides in a module enriched for contractile genes and CAD/BP GWAS hits. Our results provide evidence that *FHL5* is the most likely causal gene at the *UFL1-FHL5* locus and may impact CAD risk by regulating a network of downstream disease associated genes. These findings may contribute to our understanding of the heritable risk for multiple vascular diseases.

PrgmNr 1115 - Identifying causal protein-cardiometabolic trait relationships using whole genome sequencing

View session detail

Author Block: **G. Png**[1,2], R. Gerlini[3,4], K. Hatzikotoulas[1], A. Barysenka[1], N. W. Rayner[1], B. Rathkolb[3,4,5], J. Aguilar-Pimentel[3], J. Rozman[3,4,6], H. Fuchs[3], V. Gailus-Durner[3], E. Tsafantakis[7], M. Karaleftheri[8], G. V. Dedoussis[9], M. HrabÄ de Angelis[3,4,10], C. Becker-Pauly[11], A. Gilly[1], E. Zeggini[1,2]; [1]Inst. of Translational Genomics, Helmholtz Zentrum MÃ¼nchen â German Res. Ctr. for Environmental Hlth., Neuherberg, Germany, [2]TUM Sch. of Med., Technical Univ. of Munich and Klinikum Rechts der Isar, Munich, Germany, [3]Inst. of Experimental Genetics, German Mouse Clinic, Helmholtz Zentrum MÃ¼nchen, German Res. Ctr. for Environmental Hlth., Neuherberg, Germany, [4]German Ctr. for Diabetes Res. (DZD), Neuherberg, Germany, [5]Inst. of Molecular Animal Breeding and Biotechnology, Gene Ctr., Ludwig-Maximilians-Univ., Munich, Germany, [6]Inst. of Molecular Genetics of the Czech Academy of Sci., Czech Ctr. for Phenogenomics, Vestec, Czech Republic, [7]Anogia Med. Ctr., Anogia, Greece, [8]Echinos Med. Ctr., Echinos, Greece, [9]Harokopio Univ., Athens, Greece, [10]Chair of Experimental Genetics, TUM Sch. of Life Sci., Technische Univ. MÃ¼nchen, Freising, Germany, [11]Inst. of Biochemistry, Unit for Degradomics of the Protease Web, Univ. of Kiel, Kiel, Germany

Disclosure Block: **G. Png:** None.

Every year, an estimated 17.9 million deaths are caused by cardiovascular diseases, while more than one million die from type 2 diabetes. Cardiometabolic diseases such as these are linked to dysregulated protein levels, and understanding the genomic influence on these proteins can help to dissect the complex biology underpinning them. Here, we perform a protein quantitative trait loci (pQTL) analysis (n=2,893) of 257 serum proteins relevant to cardiometabolic processes. Meta-analysing deep whole-genome sequencing data from two Greek isolated cohorts, MANOLIS (n=1,356; 22.5x WGS) and Pomak (n=1,537; 18.4x WGS), we detect 309 independently-associated pQTL variants for 176 proteins, including 12 rare variants (minor allele frequency [MAF] cis- and three *trans*-acting) for 13 proteins that pass study-wide significance in both cohorts. Using replicating variants as instrumental variables, we find evidence of causal relationships between 35 proteins and various cardiometabolic traits via two-sample Mendelian randomisation. These include LRIG1 for atrial fibrillation and type 2 diabetes, SUMF2 for cholesterol levels, and MEP1B for high density lipoprotein (HDL) levels. Specifically, decreased MEP1B was causal for increased HDL levels. We sought *in vivo* validation in *Mep1Î²* knockout (KO) mice, where we observe significantly increased body weight and fat/lean mass ratio in female KO mice compared to their wild-type counterparts. We therefore derive causal protein-cardiometabolic trait relationships from novel pQTL data, and present orthogonal evidence supporting our findings.

PrgmNr 1116 - Network medicine-based multimodal omics discovery and iPSC-based validation of metformin for potential treatment of atrial fibrillation

View session detail

Author Block: **J. A. Castrillon**[1], Y. Zhou[1], J. Rennison[2], D. Van Wagoner[2], J. Barnard[3], F. Cheng[1], M. Chung[2]; [1]Cleveland Clinic Lerner Res. Inst., Genomic Med. Inst., Cleveland, OH, [2]Cleveland Clinic Lerner Res. Inst., Dept. of Cardiovascular and Metabolic Sci., Cleveland, OH, [3]Cleveland Clinic Lerner Res. Inst., Quantitative Hlth.Sci., Cleveland, OH
Disclosure Block: **J.A. Castrillon:** None.
Introduction: Atrial Fibrillation (AF) is the most prevalent cardiac arrhythmia in the United States, affecting 1% to 2% of the general population, significantly contributing to cardiovascular morbidity and mortality. Effective, safe, and durable therapeutic interventions are lacking. Thus, it is imperative to improve clinical management by utilizing novel approaches for drug repurposing. To date, multimodal omics analysis of genomics, transcriptomics, and interactomics (protein-protein interactions [PPIs]) have not been fully exploited for AF drug repurposing.
Hypothesis: We hypothesize that using Network Medicine-based multimodal omics analysis for drug prioritization and functional observations using induced pluripotent stem cell (iPSC) models can help us identify effective therapeutic strategies and elucidate drugâ□□s mechanisms-of-action for restoring normal sinus rhythm.
Methods: Here, we used RNA-sequencing data from a biobank of left atrial tissues obtained from cardiac surgery patients with abnormal and normal sinus rhythm to identify testable AF disease modules under the human protein-protein interactome network model. We systematically prioritized repurposable drug candidates for AF using both network proximity and gene-set enrichment analysis approaches by leveraging AF disease module findings, drug-target network, and drug-induced gene signatures in human cell lines. RNAseq was used to assess the transcriptomic impact of a top drug candidate, metformin, on gene expression of atrial myocytes differentiated from human iPSCs (a-iCMs).
Results: Via network-based screening of 2,891 FDA-approved or clinical drugs, we found nine putative drug candidates (including Metformin, Furosemide, and Rofecoxib) using both network proximity z-score and GSEA-based enrichment scores. RNA-seq analysis of metformin-treated a-iCMs identified 238 differentially expressed genes following multiple test correction (P**Conclusions:** In summary, this study presented state-of-the-art network medicine methodologies for AF drug repurposing and identified metformin as a candidate treatment for AF patients.

PrgmNr 1119 - The impact of evolutionary processes in shaping the genetics of complex traits in East Asia: A specific contribution from Denisovan introgression

View session detail

Author Block: **D. Koller**, F. R. Wendt, G. A. Pathak, A. De Lillo, F. De Angelis, B. Cabrera-Mendoza, R. Polimanti; Yale Sch. of Med., West Haven, CT
Disclosure Block: **D. Koller:** None.
Consistent evidence of how human evolution shaped the polygenicity of human traits and diseases has been observed in populations of European descent. However, no information is currently available about its impact on other ancestry groups. Here, we investigated how different evolutionary processes affected the genetics of traits and diseases in individuals of East Asian descent by leveraging genome-wide association statistics from the Biobank Japan (up to 158,284 participants). We assessed natural selection, archaic hominin introgression (*Denisova hominins* and *Homo neanderthalensis*), and several genomic functional categories regarding the heritability of physiological and pathological conditions. Similar to findings in European descent populations, the heritability of several traits was significantly enriched for annotations related to negative selection (e.g., serum creatinine: 8.8-fold enrichment for B-statistics, p=5.68Ã□10$^{â□□4}$). Regarding Denisovan introgression, we identified enrichments for left ventricular mass index (2.9-fold enrichment, p = 3.03Ã□10$^{â□□4}$), atopic dermatitis (9.2-fold enrichment, p=3.03Ã□10$^{â□□4}$), and coronary artery disease (1.7-fold enrichment, p=0.003). Neanderthal local ancestry was depleted for monocyte count (9.7-fold depletion, p = 0.002) and platelet count (4.9-fold depletion, p=0.006). To follow up these archaic-introgression enrichments, we conducted a phenome-wide association study of Denisovan- and Neanderthal alleles in participants of East Asian descent and other ancestral backgrounds from the UK Biobank. Denisovan- and Neanderthal-introgressed alleles were associated with 13 and 10 phenotypes, including metabolic, dermatological, immunological, cardiovascular, and endocrine traits. The strongest association was observed for the Denisovan-inherited locus rs59185462 with rheumatoid arthritis (beta = 0.82, p = 1.91Ã□10^{-105}). In summary, our study provides the first evidence regarding the impact of evolutionary processes on the genetics of complex traits, highlighting the specific contribution of Denisovan introgression in East Asian populations.

PrgmNr 1120 - The influence of Neanderthal introgression on human circadian biology

View session detail

Author Block: **K. Velazquez-Arcelay**[1], L. L. Colbran[2], E. McArthur[1], D. C. Rinker[1], J. A. Capra[3]; [1]Vanderbilt Univ., Nashville, TN, [2]Philadelphia, PA, [3]Univ. of California San Francisco, San Francisco, CA
Disclosure Block: **K. Velazquez-Arcelay:** None.
As the ancestors of modern Eurasians migrated out of Africa, they encountered novel environmental pressures, such as differences in day length, temperature, and pathogen exposures. These anatomically modern humans (AMH) also encountered groups of archaic humans in Eurasia, namely Neanderthals and Denisovans. For more than 400,000 years, these Eurasian hominins evolved and adapted to the variety of environments outside Africa. There is evidence of interbreeding events between these groups, resulting in the inheritance of parts of archaic genomes remaining in modern humans. This process potentially accelerated the adaptation of AMH to new environmental factors including reduced yearly ultra-violet radiation exposure and changing seasons. However, whether these groups differed substantially in circadian biology and chronotype, and whether archaic introgression contributed to the biology of human chronotype adaptation remains unknown. To address this question, we traced the evolution of chronotype based on analysis of available ancient DNA (aDNA) data from ~2,000 ancient Eurasian humans, four archaic hominins, and also genomes from thousands of present-day humans. First, we contrast circadian gene regulation patterns between archaic hominins and modern humans and identify 32 circadian genes likely divergently regulated between present-day Africans and archaic hominins, including PER2 and CLOCK. This suggests the potential for introgression to modify circadian gene regulation. Testing this hypothesis, we find that introgressed Neanderthal variants are enriched among eQTLs for circadian genes. Supporting the relevance of these gene regulatory effects on circadian traits, we found that introgressed alleles are enriched for heritability of morningness in the UK Biobank and an independent GWAS of chronotype, using stratified LD score regression. Finally, analyzing modern and ancient human DNA, we identify 1,287 circadian gene variants with significant allele frequency divergence by latitude in ancient humans. In addition to the introgressed circadian eQTL identified, we found many variants with diverse evolutionary histories with associations with traits related to chronotype. These findings demonstrate differences in circadian regulation between modern humans and archaic hominins and establish the importance of adaptive introgression on chronotype, a trait highly dependent on environmental variation.

PrgmNr 1121 - Systematic genomic investigation finds evidence of germline and somatic roles of Neanderthal introgression in liver cancer

View session detail

Author Block: **A. M. Taravella Oill**[1,2], K. H. Buetow[1,2], M. A. Wilson[1,2]; [1]Sch. of Life Sci., Arizona State Univ., Tempe, AZ, [2]Ctr. for Evolution and Med., Arizona State Univ., Tempe, AZ
Disclosure Block: **A.M. Taravella Oill:** None.
European and East Asian populations retain genetic relics of ancient contact with Neanderthals. Regions of introgressed Neanderthal DNA are often depleted in functional regions of the genome, but occasionally haplotypes remain that are predicted to provide some benefits to modern humans. In particular, there is limited understanding of the relationship between Neanderthal introgression and cancer risk. To better understand this, we have undertaken a systematic study of introgression in cancer, with a focus on liver cancer. Using publicly available introgression maps generated from European and Asian populations from the 1000 Genomes resource, we first tested whether Neanderthal introgression is enhanced or depleted in different categories of genes etiologically important in cancer among the general population. Recognizing that different cancers have variations in their underlying genetic etiology, we also examined germline and somatic changes in a single cancer - hepatocellular carcinoma - using germline and somatic variant data from 177 individuals of European ancestry and 159 individuals of East Asian ancestry from The Cancer Genome Atlas and find evidence for both risk and protective Neanderthal alleles. Among the general population, we do not find evidence for an excess or depletion of Neanderthal introgression at cancer risk genes. In liver cancer samples, we find genes with Neanderthal introgression (*LARS1* in Europeans and East Asians, *APTX* and *MCMBP* in Europeans only) exhibit an excess of somatic mutations compared to non-introgressed regions, suggesting introgressed regions may be more permissive to mutation. Provocatively, in people of East Asian descent, we find two genes (*CYB561D2* and *CACNA2D2*) that are enriched for Neanderthal introgression in the general population but depleted for introgression in people with liver cancer, suggesting introgression in these regions may be protective. As liver cancer is the second leading cause of cancer death worldwide, understanding the landscape and effects of Neanderthal introgression may inform about future susceptibility, as well as potential biomarkers for risk and therapeutic targets.

PrgmNr 1122 - Biobank-scale inference of ancestral recombination graphs reveals association to rare and common variants in complex traits

View session detail

Author Block: **B. Zhang**[1], A. Biddanda[1], P. Palamara[1,2]; [1]Dept. of Statistics, Univ. of Oxford, Oxford, United Kingdom, [2]Wellcome Ctr. for Human Genetics, Oxford, United Kingdom
Disclosure Block: **B. Zhang:** None.
The genealogical history of a set of genomes can be compactly represented using an ancestral recombination graph (ARG). We developed ARG-Needle, a method for inferring the ARG of hundreds of thousands of samples from array data, along with a mixed-model framework that leverages the ARG to detect unobserved genomic variants linked with complex traits. In extensive simulations across a variety of accuracy metrics, ARG-Needle matches or outperforms Relate (Speidel et al. Nat Genet 2019) while scaling to substantially larger sample sizes. Compared to tsinfer (Kelleher et al. Nat Genet 2019), a scalable method that only outputs ARG topology, ARG-Needle performs better on all metrics and infers both branch lengths and topology.
Using ARG-Needle, we built the genome-wide ARG for 337K unrelated European UK Biobank individuals from SNP array data. We performed mixed-model association of inferred ARG variants (focusing at first on MAF > 1%) and standing height, and compared to association of imputed variants from 65K HRC sequences (MAF > 0.1%). ARG-based association (permutation p We then tested for association between ultra-rare variants (URV, MAF These results demonstrate the impact of genealogical inference in complementing reference-based imputation for the study of complex traits and diseases, opening new possibilities for the analysis of under-sequenced cohorts for which reference panels are not yet available.

PrgmNr 1123 - Local ancestry inference for ~7,000 genomes in the Genome Aggregation Database characterizes fine-scale allele frequencies within admixed American populations

View session detail

Author Block: **M. Wilson**[1,2], K. J. Karczewski[1,2], Z. Koenig[1,2], A. R. Martin[1,2], gnomAD Consortium, M. J. Daly[2,1,3], G. Tiao[1,2], E. G. Atkinson[4,1]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Inst. for Molecular Med. Finland, Helsinki, Finland, [4]Baylor Coll. of Med., Houston, TX

Disclosure Block: **M. Wilson:** None.

Currently, there is not a comprehensive understanding of the allele frequency (AF) spectrum in Native American (NatAm) and African haplotype tracts in admixed American populations. Accurate AF estimates are critical for clinical variant interpretation. The absence of this information contributes to health disparities for admixed American populations, some of which are at greater risk for developing common diseases including cardiometabolic disorders.

Here we describe the creation of a refined population frequency reference for the Genome Aggregation Database (gnomAD) for NatAm haplotypes in 6,972 Latinx individuals. NatAm haplotypes make up a significant component (13-48%) of the complex, admixed genetic structure of Latinx individuals. To define NatAm ancestry-specific haplotypes, we constructed a continental reference panel from a tailored subset of populations within the 1000 Genomes Project and the Human Genome Diversity Project. We then used this reference panel for phasing and local ancestry inference to assign a genetic ancestry to each chromosomal segment in gnomAD Latinx individuals. Finally, we used the tool Tractor to generate local ancestry dosages at each site and compute ancestry-specific AFs. We tested multiple local ancestry inference methodologies, reference panel compositions, and phasing tools to optimize for maximal efficiency and accuracy. This unified computational pipeline was written using the Hail Batch Python module and is publicly available through GitHub [github.com/broadinstitute/gnomad_local_ancestry].

Preliminary analysis reveals a highly diverse set of samples. The average estimated proportions of global ancestry of gnomAD Latinx individuals are 30% NatAm, 12% African, and 58% European. 5% of these individuals are homogeneous, 60% are two-way admixed, and 35% are three-way admixed, with subsets of individuals most closely resembling several distinct reference populations.

This effort will result in the largest publicly available dataset for NatAm-specific AFs, refined population frequencies within the gnomAD browser, and a computational pipeline to readily resolve AFs within admixed populations. Finer AF resolution increases gnomADâ□□s clinical utility for individuals of admixed ancestry by enabling better interpretation of rare variation. The decomposition of ancestral alleles will also enable us to build a demographic model to date admixture events within these populations and determine allelesâ□□ ancestral origins. The AF spectrums and quality metrics are available for direct download and for interactive browsing on the gnomAD website, [gnomad.broadinstitute.org].

PrgmNr 1124 - Sex-specific effective population size inference in recent history

View session detail

Author Block: **R. Cai**[1], B. Browning[2], S. Browning[2]; [1]Dept Biostatistics, Univ. of Washington, Seattle, WA, [2]Univ of Washington, Seattle, WA
Disclosure Block: **R. Cai:** None.
The study of sex-specific population history is crucial for understanding if the males and females in a population have experienced differing demographic processes in the past, such as sex-specific migration processes or sex-biased family structures. Sex-specific effective population size serves as an important characterization of sex-specific population history. An approach to estimating recent effective population size is to reconstruct coalescence history through identity-by-decent (IBD) segments. Since human X chromosome and autosomes have different meiosis and recombination processes, the effective population size estimated from X chromosomal IBD segments and that from autosomal IBD segments would reflect the difference in the coalescence history of these two types of chromosomes. Therefore, by understanding the specific contribution of each sex to the coalescence history of X chromosome and autosomes, we can estimate sex-specific effective population size from the overall effective population size of X chromosome and that of autosomes.
In this study, we proposed a method to estimate sex-specific effective population size and the effective proportion of females in the population, known as the effective sex ratio (ESR), based on the distribution of IBD segment lengths and the overall effective population size of X chromosome and that of autosomes. We applied the hap-ibd program [1] and the ibd-ends program [2] to obtain the length distribution of IBD segments and estimated the overall effective population size with the IBDNe program [3]. Our method is able to correctly recover the sex-specific population size and the ESR in simulated data. Our method not only provides visualization of the sex-specific population dynamics through sex-specific population size but also estimates the ESR that could be used to test hypotheses on the sex ratio in a population. It enables us to understand sex-specific demographic dynamics that may not be revealed by the analysis of autosomes alone.
[1] Y Zhou, S R Browning, and B L Browning (2020). A fast and simple method for detecting identity by descent segments in large-scale data. The American Journal of Human Genetics 106:426-437. [2] S R Browning and B L Browning (2020). Probabilistic estimation of identity by descent segment endpoints and detection of recent selection. American Journal of Human Genetics 107:895-910.[3] S R Browning and B L Browning (2015). Accurate non-parametric estimation of recent effective population size from segments of identity by descent. The American Journal of Human Genetics 97:404-418.

PrgmNr 1127 - Local ancestry allows for improved genomic prediction in underrepresented and admixed populations

View session detail

Author Block: **E. G. Atkinson**[1], M. Kamariza[2], K. Tsuo[3], Z. Koenig[4], M. J. Daly[5], B. M. Neale[5], A. R. Martin[5]; [1]Baylor Coll. of Med., Houston, TX, [2]Harvard Univ., Boston, MA, [3]Harvard Coll., Cambridge, MA, [4]Broad Inst. of MIT and Harvard, Brighton, MA, [5]Massachusetts Gen. Hosp., Boston, MA
Disclosure Block: **E.G. Atkinson:** None.
Due to the paucity of methodological and computational approaches that account for their genomic complexity, admixed populations are systematically excluded from statistical genomic studies. Admixed populations make up more than a third of the US populace but are severely underrepresented in biomedical research which may contribute to health disparities. To reap the full benefits from the ongoing efforts to collect samples from underrepresented populations and from existing mixed ancestry cohorts, tools facilitating the well-calibrated research of admixed peoples are urgently needed.

We recently developed a local ancestry aware GWAS model, Tractor, which corrects for fine-scale population structure at the genotype level, often boosts locus discovery power, and produces ancestry-specific effect size estimates and p values. Using Tractor summary statistics from African ancestry (AFR) tracts in ~4500 admixed UK Biobank (UKB) individuals, we built polygenic risk scores (PRS) and predicted blood panel phenotypes on homogenous African ancestry UKB individuals. We benchmarked these PRS against scores created from traditional GWAS runs on 1) the same admixed cohort, 2) a large European UKB sample, and 3) a large multi-ancestry meta-analysis of continental ancestry groups from the pan-UKB project (https://pan.ukbb.broadinstitute.org/). We also tested the accuracy of several PRS models including pruning and thresholding and PRS-CSx. We find that incorporating diverse samples and ancestry-specific estimates from admixed populations results in higher prediction accuracy for homogeneous AFR individuals. The bulk of African-descent GWAS participants are currently admixed individuals of the Americas, and some underrepresented ancestries are rarely found outside of the admixed context. Thus, building models based on ancestry-specific estimates generated from the deconvolved local ancestry tracts of admixed genomes allows for better PRS performance on many diverse populations from making better use of existing collections.

We additionally highlight several loci which we find to have well-demonstrated effect size differences across ancestries, a phenomenon for which there are few prior examples in the literature. As our models are constructed off of local ancestry components from the same admixed individuals, these results hint at genetic differences rather than environmental factors, which are often tricky to disentangle. Ultimately, our work highlights how Tractor and local ancestry allow for improved population characterization and can be leveraged to advance the understanding of complex diseases across diverse cohorts.

PrgmNr 1128 - Polygenic risk prediction of obesity across the life course and in diverse populations

View session detail

Author Block: **R. A. Smit**[1,2,3,4], K. H. Wade[5], Q. Hui[6,7], M. H. Preuss[1,3,4], J. D. Arias[8], L. Yengo[9], X. Yin[10], S. E. Graham[11], M. Graff[12], K. E. North[12], E. Marouli[13], A. E. Locke[14,10], J. N. Hirschhorn[15,16,17], N. J. Timpson[5], Y. Sun[6,7], S. I. Berndt[8], R. Loos[1,3,4,18], Genetic Investigation of ANthropometric Traits (GIANT) Consortium; [1]The Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Dept. of Clinical Epidemiology, Leiden Univ. Med. Ctr., Leiden, Netherlands, [3]The Genetics of Obesity and Related Metabolic Traits Program, Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]The Mindich Child Hlth.and Dev. Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]MRC Integrative Epidemiology Unit, Univ. of Bristol, Bristol, United Kingdom, [6]Dept. of Epidemiology, Emory Univ. Rollins Sch. of Publ. Hlth., Atlanta, GA, [7]Atlanta VA Hlth.Care System, Decatur, GA, [8]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, Bethesda, MD, [9]Inst. for Molecular BioSci., The Univ. of Queensland, Brisbane, Australia, [10]Dept. of Biostatistics and Ctr. for Statistical Genetics, Univ. of Michigan Sch. of Publ. Hlth., Ann Arbor, MI, [11]Dept. of Internal Med., Div. of Cardiology, Univ. of Michigan, Ann Arbor, MI, [12]Dept. of Epidemiology, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [13]William Harvey Res. Inst., Barts and The London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom, [14]McDonnell Genome Inst., Washington Univ. Sch. of Med., St. Louis, MO, [15]Div. of Endocrinology and Ctr. for Basic and Translational Obesity Res., Boston Children's Hosp., Boston, MA, [16]Program in Med. and Population Genetics, Broad Inst., Boston, MA, [17]Dept.s of Genetics and Pediatrics, Harvard Med. Sch., Boston, MA, [18]NNF Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, Denmark

Disclosure Block: **R.A. Smit:** None.

Polygenic risk scores (PRSs) for body mass index (BMI) that leverage the increasing genome-wide association study (GWAS) sample sizes may aid risk stratification and allow targeted prevention of obesity at an early age. We constructed ancestry-specific and trans-ancestral PRSs to predict obesity in adulthood, and examined their added value over and above easily measurable predictors of obesity during childhood and adolescence.

We calculated PRSs based on summary statistics of up to 1.2 million common variants [minor allele frequency (MAF)>1%] from the GIANT consortiumâ□□s BMI GWAS meta-analysis of up to 1.6 million individuals (72% European (EUR), 16% East Asian (EAS), 6% African (AA), 4% Hispanic (HA), 2% South Asian (SAS)). Explained variance for BMI and discrimination for obesity were examined in the UK Biobank (UKB, n=437k) and Million Veteran Program (MVP, n=101k). The best performing PRS in EUR was taken forward to the Avon Longitudinal Study of Parents and Children (ALSPAC, n=5.8k), for cross-sectional and longitudinal associations with BMI across 21 time-points from birth to age 22y. We compared the predictive performance of the PRS to that of clinically available factors (maternal education, pre-pregnancy maternal BMI, household social status).

The trans-ancestral PRS explained more of the variation in BMI than ancestry-specific PRSs in all but the EUR-populations (R^2 min-max for non-EUR; UKB: 7.5-12.4% (AA/EAS); MVP: 5.7-11.1% (AA/HA); UKB-EUR (EUR-PRS): 15.8%; MVP-EUR (EUR-PRS): 13.1%). For all ancestries, maximum explained variance was roughly double that of previously published obesity PRSs. The PRSs were better at discriminating between adults with or without obesity than age, sex, or scores of genome-wide significant variants only. EUR-PRS associations with BMI were weak at birth, but increased rapidly during childhood, and remained stable from adolescence onwards (e.g., BMI-SD per PRS-SD at 13y (95%CI): 0.39 (0.37,0.42)). Consistently, longitudinal modeling of BMI trajectories using the PRS showed increasing divergence until early adolescence. When added to other factors available at birth, the PRS helped predict substantially more of BMI from 5y onwards (e.g., R^2 at 5y: 13 to 18%; 11y: 11

to 21%).
The current PRSs, based on larger GWAS sample sizes, double the previously explained variance for BMI across multiple ancestries, thereby advancing the options for prognostication in populations traditionally underrepresented in genetic research. Moreover, we find that genetic predisposition to adult obesity affects childhood growth trajectories, and shows potential to improve risk stratification for obesity at an early age.

PrgmNr 1129 - Improving Polygenic Prediction in Ancestrally Diverse Populations

View session detail

Author Block: **Y. Ruan**[1,2], Y-C. Feng[1,3,4,5], C-Y. Chen[6], M. Lam[1,5,7,8,9], Stanley Global Asia Initiatives, A. Sawa[10,11], A. R. Martin[1,5], S. Qin[2], H. Huang[1,5,12], T. Ge[1,3,4]; [1]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Bio-X Inst.s, Key Lab. for the Genetics of Dev.al and Neuropsychiatric Disorders (Ministry of Ed.), Shanghai Jiao Tong Univ., Shanghai, China, [3]Dept. of Psychiatry, Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA, [4]Psychiatric and Neurodevelopmental Genetics Unit, Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [5]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [6]Biogen, Cambridge, MA, [7]Div. of Psychiatry Res., The Zucker Hillside Hosp., Northwell Hlth., Glen Oaks, NY, [8]Res. Div., Inst. of Mental Hlth.Singapore, Singapore, Singapore, [9]Human Genetics, Genome Inst. of Singapore, Singapore, Singapore, [10]Dept.s of Psychiatry, NeuroSci., and BioMed. Engineering, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [11]Dept. of Mental Hlth., Johns Hopkins Univ. Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [12]Dept. of Med., Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA
Disclosure Block: **Y. Ruan:** None.

Polygenic risk scores (PRS) are less effective when ported across populations. While the scale of non-European genomic resources has been expanded in recent years, a clear attenuation of the predictive performance of PRS remains in individuals who are genetically distant from Europeans.

In order to include data from all ancestral groups to ensure more equitable delivery of genomic prediction to global populations, we developed the first principled Bayesian PRS construction method, termed PRS-CSx, that jointly models GWAS summary statistics from multiple populations to improve cross-ancestry polygenic prediction. PRS-CSx couples genetic effects across populations via a shared continuous shrinkage prior, enabling more accurate effect size estimation by sharing information between summary statistics and leveraging linkage disequilibrium (LD) diversity across discovery samples, while inheriting computational efficiency and robustness from PRS-CS.

PRS-CSx outperformed existing PRS methods across various simulations settings with different sample sizes, fractions of causal variants, and genetic correlations between populations. Using quantitative traits from biobanks, we showed that PRS-CSx substantially improved the prediction accuracy even if only a small non-European GWAS was included in the discovery data. For example, the median R2 increased by 76% for individuals of East Asian ancestry when the Biobank Japan samples (N=62K-159K) were added to the UK Biobank European samples (N=340K-360K) to train the PRS. Similarly, the median R2 increased by 22% for individuals of African ancestry when the PAGE study samples (N=20K-50K) were integrated with UK Biobank and Biobank Japan samples (400K-519K). Furthermore, by integrating GWAS summary statistics of schizophrenia from East Asian (14K-17K cases due to leave-one-out) and European (33K cases) populations, PRS-CSx more accurately predicted schizophrenia risk in individuals of East Asian ancestry, showing 52% and 97% improvement in the liability R2 relative to PRS constructed using East Asian or European summary statistics only, and approximately doubled the prediction accuracy when compared with alternative methods that can combine multiple GWAS to make prediction.

Our method represents a much needed and critical breakthrough in PRS construction. Through joint modeling of multi-ancestry data, PRS-CSx substantially improves polygenic prediction in non-European populations. With the rapid expansion of non-European genomic resources, our method will help accelerate the equitable deployment of PRS in clinical settings and maximize its healthcare potential.

PrgmNr 1130 - A trans-ancestry polygenic test to predict severe hypercholesterolemia in diverse ancestry patients

View session detail

Author Block: **M. C. Turchin**[1], R. A. J. Smit[2,3], S. W. Choi[4], G. M. Belbin[1,5], C. Hoggart[4], M. Preuss[2,3], M. Guindo-Martinez[6], Z. Wang[2,3], B. Murphy[5], J. H. Cho[2,4,5], R. Do[2,5], R. J. F. Loos[2,3,6], P. F. O'Reilly[4], N. S. Abul-Husn[1,4,5], E. E. Kenny[1,4,5]; [1]The Inst. for Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]The Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]The Mindich Child Hlth.and Dev. Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Dept. of Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]The Novo Nordisk Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, Denmark

Disclosure Block: **M.C. Turchin:** None.

Approximately 7% of adults have severe hypercholesterolemia (SH; untreated low density lipoprotein (LDL-C) â¥ 190 mg/dL). SH is associated with a 6-fold increased risk of cardiovascular disease, and up to 20-fold increased risk in individuals identified with monogenic Familial Hypercholesterolemia (FH)-associated variants. Despite high frequency of cholesterol screening and awareness, individuals with SH remain undertreated, with disparities in treatment and LDL-C control observed among African American (AA) populations. Only 2.5% of individuals with SH harbor a monogenic FH-associated variant, and polygenic SH accounts for 15%-30% of clinical FH, motivating the development of a polygenic test for predicting SH in diverse populations. We obtained summary statistics for validated trans-ancestry polygenic risk scores (PRS) to predict LDL-C from the Global Lipids Genetics Consortium pre-publication. The PRS were developed from a genome-wide association study of ~1.6M trans-ethnic participants, and validated in European (EU), AA, African, Hispanic or Latino (HL), South Asian and East Asian populations. We leveraged independent genotype and phenotype data from the diverse BioMe biobank in New York City. We extracted laboratory values and medications from electronic health records for adults with an age range of 18-95 from three population groups: AA, EU, and HL (other groups were excluded due to low sample size). SH cases were defined as participants with statin-adjusted maximum LDL-C â¥ 190 mg/dL and controls with statin-adjusted maximum LDL-C

PrgmNr 1131 - Phenome-wide association study of polygenic risk for asthma in the UK Biobank highlights traits with shared genetic architecture and sex specific effects

View session detail

Author Block: **Y. Lee**[1], W. Wentworth-Sheilds[2], H. Im[3], C. Ober[3], N. Schoettler[3], M. Dapas[2]; [1]Biological Sci. Collegiate Div., Univ Chicago, Chicago, IL, [2]Dept Human Gen, Univ Chicago, Chicago, IL, [3]Univ Chicago, Chicago, IL
Disclosure Block: **Y. Lee:** None.
Polygenic risk scores (PRSs) aggregate additive effects of genetic variants to estimate individual risks for heritable diseases and can be used clinically to inform decisions on screening, therapeutic intervention, and lifestyle modification. The aim of this study was to develop a PRS for asthma using genetic information from a large, multiethnic (ME) cohort and investigate its association with 267 phenotypes in the UK Biobank (UKB). Two asthma PRS models were developed based on European (EU) (19,954 cases, 107,715 controls) and ME (23,948 cases, 118,538 controls) summary statistics from the Trans-National Asthma Genetic Consortium meta-analysis. Posterior SNP effect size estimates were generated using a Bayesian regression framework, implemented in PRS-CS. To evaluate PRS prediction for asthma, each model was applied to white British (36,065 cases, 314,781 controls) and ME (43,109 cases, 377,061 controls) subjects from UKB using logistic regressions adjusting for sex and ancestry. The EU PRS applied to the white British cohort had the strongest association with doctor-diagnosed asthma ($p=1.96x10^{-295}$, OR=1.34, 95% CI=1.32-1.35, AUC=0.582) and was most strongly associated with childhood onset asthma (COA; onset before age 12; $p=1.77x10^{-181}$, OR=1.59, 95% CI=1.54-1.64, AUC=0.624). There were significant sex-by-PRS interaction effects for COA (p=0.049) and adult onset asthma (AOA; onset after age 25; p=0.048). Given the same PRS, males had a higher risk than females for COA but females had a higher risk than males for AOA. The phenome-wide association study identified significant associations between the PRS and 27 binary and 69 quantitative traits (Bonferroni p-4). The most significant association was with percent eosinophils ($p=9.33x10^{-298}$, $\hat{I}^2=0.11$), a known asthma-associated trait. Other associated traits included asthma age of onset ($p=4.12x10^{-94}$) and measures of lung function (FEV_1) ($p=1.91x10^{-117}$). Some associations were less expected. For example, age at first live birth was negatively correlated with the PRS ($p=8.57x10^{-15}$, $\hat{I}^2=-0.095$) and HbA1c was positively correlated with the PRS ($p=4.83x10^{-33}$, $\hat{I}^2=0.13$). Sex-specific effects were observed for 5 binary and 15 quantitative traits, such as fat-free mass ($p=1.71x10^{-6}$, $\hat{I}^2=0.028$ in females; p=0.42, $\hat{I}^2=7.6x10^{-3}$ in males). Overall, our results suggest shared genetic architectures between asthma and a broad swath of pulmonary, cardiometabolic, anthropometric, and reproductive traits, many of which had not previously been linked to asthma and some with sex-specific effects. This research was conducted using the UK Biobank Resource under application number 44300.

PrgmNr 1132 - Modelling hidden genetic risk from family history for improved polygenic risk prediction

View session detail

Author Block: **T. Lu**[1], V. Forgetta[2], B. Richards[1], C. M. Greenwood[3]; [1]McGill Univ., Montreal, QC, Canada, [2]Lady Davis Inst., Montreal, QC, Canada, [3]Jewish Gen. Hosp., Montreal, QC, Canada
Disclosure Block: **T. Lu:** None.
With many polygenic risk scores demonstrating research and clinical utility, it is worth questioning whether family history, a traditional genetic predictor, still provides valuable information.
Family history of complex traits may be influenced by transmitted rare pathogenic variants, intra-familial shared exposures to environmental factors, as well as a common genetic predisposition. Therefore, we propose and develop a latent factor model to quantify disease risk in excess of that captured by a common SNP-based polygenic risk score, but inferable from family history. This model enables calibration of polygenic risk scores with respect to family history without fitting regression models.
We applied our model to predict adult height for 941 children in the Avon Longitudinal Study of Parents and Children. Our predictor was able to explain ~55% of the total variance in adult height, close to the estimated heritability of height and substantially higher than ~40% captured by a polygenic risk score for height or mid-parental height alone. For nine complex diseases, including metabolic syndromes, cardiovascular diseases, neurological disorders and several types of cancer, we used our model to improve polygenic risk prediction for >400,000 White British participants in the UK Biobank. For all nine complex diseases investigated in the UK Biobank, parental disease history brought significant improvements in the discriminative power of polygenic risk prediction. For instance, combined with age and sex, our predictor achieved an area under the receiver operating characteristic curve (AUROC) of 0.734 and an area under the precision-recall curve (AUPRC) of 0.171 in identifying individuals with type 2 diabetes, exhibiting significantly stronger discriminative power than the polygenic risk score (AUROC = 0.712; AUPRC = 0.148) or the parental disease history (AUROC = 0.707; AUPRC = 0.148) alone. Comparing to using a type 2 diabetes polygenic risk score, our predictor had a net reclassification index of 3.72% in identifying 20% of the population at an elevated risk.
Taken together, our work showcases an innovative paradigm for risk calculation, and supports the utility of incorporating family history into polygenic risk score-based genetic risk prediction models.

PrgmNr 1135 - A genome-wide mutational constraint map quantified from variation in 76,156 human genomes

View session detail

Author Block: **S. Chen**[1], R. Collins[1], Q. Wang[2], D. G. MacArthur[3], M. E. Talkowski[1], gnomAD Project Consortium, M. J. Daly[1], B. M. Neale[1], K. J. Karczewski[4]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Garvan Inst., Darlinghurst, Australia, [4]Broad Inst., Medford, MA
Disclosure Block: **S. Chen:** None.
Assigning function to non-coding variation is a significant challenge in human genomics. While intolerance to protein-coding variation is a powerful predictor of the phenotypic consequences of gene disruption, attempts to assess intolerance in non-coding regions have proven more difficult. This is largely due to the lack of genetic code in non-coding DNA for predictable variant effects and the relatively small sample sizes of WGS reference data. Earlier studies have assessed non-coding function through evolutionary conservation, but these do not necessarily capture human-specific constraint.

Here we leverage 76,156 human genomes from the Genome Aggregation Database (gnomAD), the largest public dataset of genomic variants, to build a map of mutational constraint for the whole genome. We present a refined mutational model, which incorporates trinucleotide sequence context, base-level methylation, and regional genomic features to predict expected levels of variation under neutrality. We trained our model on 450K *de novo* variants previously detected in 5.5K parent-child pairs and applied it to assess constraint across the entire genome. Specifically, we divided the genome into non-overlapping 1kb windows and quantified constraint for each window by comparing the expected variation and the observed variation in the gnomAD dataset (totaling 400M high-quality rare variants). As expected, protein-coding sequences were skewed towards the most constrained percentile (Z>4) of the spectrum (2.7-fold compared to unconstrained Z-200). We further computed constraint against structural variation and found a strong correlation (p-200), internally validating our approach.

Examining exclusively on non-coding regions revealed significant enrichment of constraint (p-30; odds ratios 1.7-6.1) for regulatory elements, evolutionarily conserved sequences, GWAS loci, and predicted deleterious variants. Remarkably, the 9p21 locus - a widely recognized and replicated risk factor for coronary artery disease - although devoid of genes, appeared to be more constrained than regions encompassing coding sequences (4-fold, p=0.07). Other examples include an intergenic variant associated with type I diabetes (T1D; rs11594656) found in the 5th most constrained per-mille of the genome, which has been shown to influence transcription at *IL2RA* involved in T1D susceptibility.

This genome-wide constraint map, quantified from the largest public WGS reference cohort, is an asset to facilitate identification and interpretation of the functional fraction of non-coding DNA in the human genome.

PrgmNr 1136 - Assessment of mitochondrial genome constraint in the gnomAD database of >50,000 individuals using a composite likelihood model

View session detail

Author Block: **N. J. Lake**[1,2], W. Liu[3], K. M. Laricchia[4], G. Tiao[5], H. Zhao[6], S. Calvo[7], S. Sunyaev[8], M. Lek[9]; [1]Dept. of Genetics, Yale Univ. Sch. of Med., New Haven, CT, [2]Murdoch Children's Res. Inst., Royal Children's Hosp., Melbourne, Australia, [3]Yale Univ., New Haven, CT, [4]Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Broad Inst MIT & Harvard, Cambridge, MA, [6]Yale Univ. Sch. of Publ. Hlth., New Haven, CT, [7]Broad Inst., Cambridge, MA, [8]Harvard Med. Sch. and Brigham & Women's Hosp., Boston, MA, [9]Yale Univ. Sch. of Med., New Haven, CT

Disclosure Block: **N.J. Lake:** None.

Mitochondria carry a ~16kb genome (mtDNA) encoding proteins, transfer RNAs and ribosomal RNAs. Pathogenic variants in mtDNA cause a range of phenotypes collectively known as mitochondrial diseases. Mitochondrial variant interpretation is challenging, and most candidate variants in individuals with disease are of uncertain significance. Constraint models for the nuclear genome have been extremely useful for the analysis and interpretation of variants. By comparing observed and expected levels of variation in a population, these models identify which regions and genes are most likely to harbor disease-causing variants. However an equivalent model for the mtDNA had not been developed, in part due to the unique features of this genome.

Here we developed the first constraint model specifically for mtDNA and applied it to the new gnomAD mtDNA database which reports variation in 56,434 individuals. We used a composite likelihood model to quantify mutability in trinucleotide contexts, as it is suited for possible sparsity of counts per context in mtDNA, and leverage de novo mutation datasets to build the model. Given selection occurs against both the number and heteroplasmy of mtDNA variants (fraction of mtDNA copies with the mutation), we applied the model to determine the expected sum of the maximum heteroplasmy of variants in genes/regions in gnomAD. We assess constraint as observed:expected (obs:exp) ratios and provide 90% confidence intervals (CI).

The model accurately predicted the level of neutral variation in gnomAD, establishing its utility (correlation coefficient R=1). We calculated the obs:exp ratio of variant types across and within each gene, and showed that known pathogenic variation was substantially constrained (eg. ClinVar ratio=0.10, CI=0.07-0.14). The overall synonymous obs:exp ratio was 1.0 (CI=0.99-1.02) and consistent across genes (CI upper bounds 0.92-1.17), unlike missense variants which showed lower and variable constraint (ratio=0.25, CI upper bounds 0.15-0.97). For tRNAs, we utilized their common structure to also characterize site-level constraint. Notably, we identified constraint in regions typically overlooked in disease analyses, such as rRNA and non-coding elements (eg. L-strand origin ratio=0.25, CI=0.18-0.32). Lastly, we demonstrate the utility of this model for variant interpretation in patients, including by showing enrichment of pathogenic variants in the most constrained regions.

In sum, this model represents a new tool for mtDNA analysis and variant interpretation. It further provides novel insight into which mtDNA regions are most important to function.

PrgmNr 1137 - Quantifying negative selection on synonymous variants

View session detail

Author Block: **M. Gudkov**, E. Giannoulatou, L. Thibaut; Victor Chang Cardiac Res. Inst., Sydney, Australia
Disclosure Block: **M. Gudkov:** None.
Most disease sequencing studies tend to focus primarily on missense and potential loss-of-function variants, such as stop-gain and frameshift variants, at the expense of other classes of mutations. In particular, synonymous genetic variants, that is, those single-nucleotide variants (SNVs) that do not alter the produced amino acid sequence, are routinely considered to be non-deleterious. However, the role of these so-called â□□silent mutationsâ□□ is potentially more important than was previously thought. For instance, synonymous SNVs (sSNVs) may create nonoptimal codons, thus affecting the stability of the produced mRNA and the overall translational efficiency.
It has also been shown that optimality-reducing sSNVs undergo purifying selection, the extent of which, nonetheless, remains unknown. The latter presents a significant limitation for variant prioritisation and, consequently, for finding the true causes of genetic disorders. Indeed, SNVs with unknown, unquantified deleteriousness are generally more likely to be overlooked or even excluded from further analysis. To fully understand the role of sSNVs in human disease, a quantitative framework is needed to assess their potential effect in comparison with other, better studied classes of mutations.
Here we quantify the intensity of the negative selection acting on all possible sSNVs. We calculated the mutability-adjusted proportion of singletons (MAPS), a recently developed metric of deleteriousness, for a QC compliant subset of sSNVs from 125,748 gnomAD exome sequences (release 2.1.1). Using MAPS, we found that optimality-reducing sSNVs are subject to stronger selection than optimality-increasing ones, which was confirmed using conservation and site-frequency spectrum analyses. Furthermore, we found that purifying selection affects sSNVs in a gene- and amino acid-dependent manner, with glutamine being particularly intolerant to such mutations. We also propose an improved version of MAPS for sSNVs.

PrgmNr 1138 - Objective targeting of *de novo* loss of function variants in constrained genes improves diagnostic rates in the 100,000 Genomes Project

View session detail

Author Block: **E. G. Seaby**[1,2,3,4], H. Brittain[5], A. Taylor Tavares[5], Genomics England Research Consortium, D. Baralle[2], H. L. Rehm[3,1], A. O'Donnell-Luria[1,3,4], S. Ennis[6]; [1]Broad Inst., Cambridge, MA, [2]Univ. of Southampton, Faculty of Med., Southampton, United Kingdom, [3]Massachusetts Gen. Hosp., Boston, MA, [4]Boston Children's Hosp., Boston, MA, [5]Genomics England, London, United Kingdom, [6]Univ. of Southampton, Southampton, United Kingdom

Disclosure Block: **E.G. Seaby:** None.

Large-scale sequencing programs are aiding in the affordability and accessibility of genome sequencing (GS) for diagnostic purposes. The 100,000 Genomes Project (100KGP) has led the way in integrating GS within a national health service; however, sequencing thousands of individuals necessitates time and resources to generate patient reports. Consequently, virtual gene panels have been applied for more efficient analysis, restricting the number of variants assessed for diagnostic reporting. Panels rely on clearly characterized phenotypes and accurate gene selection, yet risk missing pathogenic variants outside of the panel applied. We propose an approach targeted at large-scale sequencing projects that objectively filters for high-yield variants in constrained genes. The LOEUF score is a continuous metric of loss-of-function gene constraint, with low scores highly correlated with haploinsufficient genes. We filtered 25,050 rare disease trios in 100KGP for *de novo*, loss-of-function variants in genes with a LOEUF score

PrgmNr 1139 - Biallelic noncoding regulatory mutations in autism spectrum disorder

View session detail

Author Block: **R. N. Doan**, J. Scotellaro, L. Kelley, T. W. Yu, C. A. Walsh; Boston Children's Hosp., Boston, MA
Disclosure Block: **R.N. Doan:** None.
The phenotypic and genetic heterogeneity of autism spectrum disorder (ASD) in children has posed immense clinical and research challenges to understanding the underlying genetic determinants of risk. Despite the association of more than 1,000 genes with ASD, many families lack an identifiable *de novo* or inherited protein altering mutation, suggesting that much of the genetic etiology lies within noncoding portions of the genome. While recent studies have begun to shed light on the role of noncoding mutations (*e.g*., gene promoters and human-specific enhancers) on ASD risk, the ability to distinguish damaging regulatory mutations from the vast numbers of benign events remains challenging. To overcome this challenge, we developed a custom variant classification system for assessing the transcriptional and translational impacts of mutations from whole genome (WGS) and whole-exome (WES) sequencing data from more than 5,000 families with ASD. Furthermore, we developed a custom-curated database consisting of more than 26,000 conserved neurally active promoter elements (CNPs; core promoters, proximal promoters, and untranslated regions [UTRs]) and regulatory regions to identify alleles with the greatest likelihood of impacting neural development. Strikingly, we find that biallelic mutations within these elements contribute to risk in at least 3% of families with ASD. Predicted damaging promoter mutations were highly enriched for established associations to ASD and intellectual disability, including more than 20 known ASD genes from SFARI gene database. Surprisingly, we identified 3 promoter-altering mutations in the *FMR1* gene, the well-established cause of Fragile X, with 2 of the mutations impacting core promoter elements (*e.g*. GC-Box) and a third mutation affecting the Kozak sequence. Beyond well-recognized genes, we identify strong candidate mutations in lesser-established ASD genes, such as a promoter mutation within*TTI2*, a gene linked to ID and ASD features, that causes the loss of an essential TFIIB-recognition element (BRE). Collectively, our data provide the first evidence that biallelic mutations impacting neurally active gene promoters and UTRs contribute to ASD risk and give insight into one possible mechanism underlying the wide phenotypic spectrum observed amongst families.

PrgmNr 1140 - Estimating the clinical diagnostic yield of whole-genome sequencing in 825 orofacial cleft trios

View session detail

Author Block: **K. K. Diaz Perez**[1], B. A. Carter[2], M. R. Bishop[1], L. C. Valencia-Ramirez[3], C. Restrepo[3], L. M. Moreno[4], T. H. Beaty[5], J. C. Murray[6], E. Feingold[7], M. L. Marazita[8], H. Brand[9], E. J. Leslie[1]; [1]Emory Univ., Atlanta, GA, [2]Agnes Scott Coll., Decatur, GA, [3]FundaciÃ³n Clinica Noel, Medellin, Colombia, [4]Univ Iowa, Iowa City, IA, [5]Johns Hopkins Univ, Sch PubHlth, Baltimore, MD, [6]U of Iowa, Iowa City, IA, [7]Univ of Pittsburgh, Pittsburgh, PA, [8]Univ Pittsburgh, Pittsburgh, PA, [9]Boston, MA
Disclosure Block: **K.K. Diaz Perez:** None.
Orofacial clefts (OFCs) are among the most common craniofacial birth defects and require years of treatment, presenting a societal and personal burden. Most OFCs occur as isolated events (nonsyndromic) and are thought to be etiologically complex, with genetic and environmental risk factors. However, common variants only account for 25% of the estimated heritability of OFCs, while whole-exome and -genome sequencing (WGS) studies increasingly support a role for rare variants with rare mutational drivers acting in a Mendelian manner in ~10% of nonsyndromic OFC cases. In support of this, we previously found a significant enrichment of *de novo* mutations (DNMs) in genes associated with autosomal dominant OFC syndromes. This study aims to comprehensively interrogate DNMs and rare inherited single-nucleotide and structural variants in a collection of 825 OFC trios to estimate the clinical diagnostic yield for sequencing in OFCs. We constructed a gene list of 510 OFC-related genes from commercially available clinical gene panels, OMIM, and expert-curated gene lists to focus our analysis. After filtering for allele frequency and variant effect, we categorized variants into a ranked tier system based on *in silico* prediction scores, gene constraint, and ClinVar annotations to prioritize potential pathogenic variants. Variants in highly ranked tiers were reviewed for possible pathogenicity using American College of Medical Genetics & Genomics (ACMG) criteria. Among OFC cases, 82% had at least one variant qualifying for ACMG review, demonstrating the difficulty in assessing the impact of rare variants by bioinformatic tools alone. We further filtered these results by focusing on protein-altering (missense and loss-of-function (LoF)) DNMs and inherited LoF variants, which are easier to interpret. We identified 46 DNMs in 7.8% of genes (5.5% cases). These mutations were in several well-established OFC genes, including *TFAP2A and IRF6*, each with 3 DNMs. We further identified 99 inherited LoF variants for review in 64 genes, accounting for 95 cases (12%). Five LoF variants were inherited from an affected parent. Moreover, various genes associated with OFCs, including *ARHGAP29* and *COL2A1*, showed an enrichment of inherited LoF mutations in our sample, each containing â¥ 3 LoF variants across all cases, highlighting the genetic heterogeneity of OFCs. Overall, using DNMs and inherited LoF variants, we conservatively estimate the diagnostic yield of WGS in OFCs to be only 5-15%. Given the complexity of OFCs, this is likely to be an underestimate but suggests clinical genetic sequencing could effectively refine recurrence risk estimates and reduce the diagnostic gap in OFCs.

PrgmNr 1143 - Modeling tissue co-regulation to quantify tissue-specific contributions to disease heritability

View session detail

Author Block: **T. Amariuta**, K. M. Siewert-Rocks, A. L. Price; Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA
Disclosure Block: **T. Amariuta:** None.
Despite abundant evidence of disease etiologies that span multiple tissues, quantifying tissue-specific contributions to disease heritability remains challenging. Previous work emphasizes the potential of accounting for tissue co-regulation (Ongen et al. 2017 Nat Genet), but tissue-specific disease effects have not been formally modeled.
We introduce a new method, tissue co-regulation score regression (TCSC), that quantifies tissue-specific contributions to disease heritability by regressing transcriptome-wide association study (TWAS) gene-disease chi-square statistics on tissue co-regulation scores, across genes and tissues. TWAS statistics include direct effects of predicted cis-genetic components of gene expression on disease and tagging effects of co-regulated tissues (Wainberg et al. 2019 Nat Genet). TCSC distinguishes causal versus tagging gene-disease effects across tissues by modeling pairwise correlations of predicted gene expression between tissues (tissue co-regulation scores). TCSC also quantifies tissue-specific contributions to genetic correlations between diseases by regressing products of TWAS z-scores from two diseases on tissue co-regulation scores. In simulations, TCSC detects causal tissues with well-calibrated false positive rate across a broad range of parameter settings. At default settings, TCSC attained 60% power to detect causal tissues, which is substantially higher than the power of the Ongen et al. method. TCSC also estimates the proportion of SNP-heritability explained by each tissue; estimates are conservative, as they exclude effects of genes with non-significant gene expression heritability at finite sample size.
We applied TCSC to 23 independent diseases and complex traits from UK Biobank (average N = 396K), using gene expression prediction models constructed from GTEx data across 47 tissues to compute TWAS statistics and co-regulation scores. Below, we discuss results for three anthropometric traits: body mass index (BMI), height, and waist-hip-ratio adjusted for BMI (WHR). For BMI, brain cerebellum was the only significant tissue after correcting for the number of tissues tested (proportion of SNP-heritability explained: 0.096 Â± 0.021); this is consistent with the known role of the central nervous system in BMI (Locke et al. 2015 Nature). For height, brain cerebellum was the most significant tissue (0.091 Â± 0.023). For WHR, skin fibroblast was the only significant tissue (0.116 Â± 0.038), consistent with the known role of connective tissue (Finucane et al. 2018 Nat Genet). In conclusion, TCSC is a powerful method for quantifying tissue-specific contributions to disease heritability.

PrgmNr 1144 - Uncovering context-specific genetic regulation of gene expression from single-cell RNA-sequencing using latent-factor models

View session detail

Author Block: **B. J. Strober**[1], K. Tayeb[2], J. M. Popp[1], M. Gordon[3], R. Perez[3], C. Ye[3], A. Battle[1]; [1]Johns Hopkins Univ, Baltimore, MD, [2]Univ. of Chicago, Chicago, IL, [3]UCSF, San Francisco, CA
Disclosure Block: **B.J. Strober:** None.
Identification of genetic variants associated with gene expression, or expression quantitative trait loci (eQTLs), can be used to better understand the regulatory mechanisms linking genetic variation with disease. However, genetic regulation of gene expression is a complex process, with genetic effects known to vary across contexts such as developmental time points, cell types, and environmental conditions. It is therefore critical to identify eQTLs from diverse contexts in order to properly characterize the molecular mechanisms underlying disease associated loci. Indeed, eQTLs from adult bulk tissue samples fail to explain the majority of known disease loci. Recent work has shown single-cell RNA-sequencing (scRNA-seq) provides unique data to uncover context-specific eQTLs; such higher-resolution data will naturally span diverse cell types and cellular states, many of which would not be observable from bulk RNA-seq. However, the relevant factors, such as cell type or state, that actually modulate genetic effects may not be known a priori. Furthermore, an individual cell may be defined by multiple, overlapping contexts, such as a particular cell type and perturbation response affecting partially overlapping sets of cells. Therefore, we developed SURGE, a novel probabilistic model that uses matrix factorization to jointly learn a continuous representation of the cellular contexts defining each measurement, and the corresponding eQTL effect sizes specific to each learned context, allowing for discovery of context-specific eQTLs without pre-specifying subsets of cells or samples. In a proof of concept using bulk expression data over 49 tissues from the GTEx project, SURGE automatically learns factors capturing tissue and cell type composition differences, in addition to one factor reflecting individual ancestry. We applied SURGE to a single-cell eQTL data set consisting of multiplexed single-cell RNA-sequencing data from over 750,000 peripheral blood mononuclear cells from 119 individuals. SURGE automatically identifies cell-type specific eQTLs from this data, identifying factors capturing continuous representations of distinct blood cell types and grouping biologically related cell types into the same factor. In summary, we provide a novel approach to automatically uncover cell types and contexts that modulate genetic regulation of gene expression, enabling the unbiased discovery of diverse context-specific eQTLs from single cell, time course, and multi-condition data, and expanding our ability to explain mechanisms underlying disease-associated loci.

PrgmNr 1145 - Acetylated chromatin domains link chromosomal organization to cell and circuit level dysfunction in schizophrenia and bipolar disorder

View session detail

Author Block: **K. Girdhar**[1], J. Bendl[1], S. Rahman[1], P. Dong[1], W. Liao[2], L. Brown[3], O. Devillers[1], B. Kassim[1], J. R. Wiseman[1], R. Park[1], E. Zharovsky[1], A. Kozlenkov[4], R. Jacobov[1], S. Dracheva[5], Y. Jiang[6], J. F. Fullard[1], S. Akbarian[1], P. Roussos[1]; [1]Mount Sinai, New York, NY, [2]New York Genome Ctr., New York, NY, [3]Yale Univ., New Haven, CT, [4]Dept. of Psychiatry, Icahn Sch. of Med. at Mount Sinai, New York, NY 10029, USA, New York, NY, [5]Mental Illness Res. Ed. and Clinical Ctr. (MIRECC), James J. Peters VA Med. Ctr., Bronx, New York, 10468, USA, New York, NY, [6]Inst. of BioMed. Sci., Fudan Univ.,, Shanghai, China
Disclosure Block: **K. Girdhar:** None.

Acetylated chromatin domains link chromosomal organization to cell and circuit level dysfunction in schizophrenia and bipolar disorder

The three-dimensional (3D) genome organization is crucial for the interplay between cell specific gene expression, chromatin structure, and other epigenetic factors that underlies our brain functions. Dysregulation of transcription and chromatin structure leads to debilitating disorders such as schizophrenia (SCZ). To explore the dysgreulation in chromatin structure in SCZ and bipolar disorder (BD), we capitalized on histone modification landscapes as they are tightly linked to chromatin structures. We mapped **739 libraries** of active promoter and enhancer-associated histone methylation (H3K4me3) and acetylation (H3K27ac) profiles in PFC from **564 brain donors**, providing to date the largest histone modification dataset for SCZ and BD. In the first part of the study, dysregulation of enhancer-associated histone peaks (H3K27ac) across SCZ (BD) cases and controls showed stronger enrichment of SCZ (BD) GWAS risk variants in enhancers than in promoters specific regions. Additionally, we noticed enrichment in SCZ variants was coming from hyperacetylated peaks but not from hypoacetylated peaks. In the second part of the study, we built chromatin structure by leveraging the inter-individual correlations between histone peaks to identify domains of physically interacting regulatory elements called cis-regulatory domains (CRDs). Our in-silico biological validation of CRDs showed tight link to the structures of self-interacting domains, with enrichments of CTCF structural proteins at CRD domain boundaries, which is in line with CTCF enrichment of PFC NeuN+ Hi-C defined TAD boundaries. Next, we investigated dysregulation in chromatin structure using CRDs expression as a metric. Interestingly, SCZ and BD sensitive CRDs which were hyperacetylated were strongly enriched for SCZ (BD) GWAS risk variants than hypoacetylated CRDs. To further explore the role of disease sensitive CRDs, we created another layer of chromatin structure that manifested A/B compartments by taking the inter-individual correlation of diseased CRDs. K-means clustering of correlation of disease-sensitive H3K27ac CRDs revealed 3 Clusters. Out of which cluster 3 with majority of hyperacetylated CRDs were strongly enriched for glutamatergic specific H3K27ac peaks and chromHMM active chromatin states from fetal brain linking the connection of SCZ and BD with neurodevelopmental mechanism.In summary, our analysis strongly hint at a chromatin structure brain pathology representative of the broader population of subjects diagnosed with SCZ and BD.

PrgmNr 1146 - Genetic determinants of alternative splicing in stimulated iPSC-derived macrophages enhance the understanding of immune-mediated disease risk

View session detail

Author Block: **O. El Garwany**, N. I. Panousis, A. Knights, N. Kumasaka, M. Imaz, A. Barnett, L. B. Vilarino, A. Tsingene, C. Gomez, D. Gaffney, C. Anderson; Wellcome Sanger Inst., Hinxton, United Kingdom

Macrophages represent the first line of defence against numerous pathogens due to their ability to respond to environmental cues. Genetic variation can lead to substantial alterations of this response, increasing susceptibility to immune-mediated diseases (IMD). Thus, understanding how macrophage cellular phenotypes such as gene expression are genetically controlled, particularly in response to environmental stimuli, can provide much needed insight into IMD biology and risk.
Since modelling environmental contexts in primary cells at scale is not practically feasible, induced Pluripotent Stem Cells (iPSC) offer an alternative *in vitro* system to model environmental contexts, whilst providing enough cells to study the genetic determinants of gene expression and its different layers of control. Alternative splicing (AS) is one of the major post-transcriptional layers of gene expression control that results in remarkable diversification of the transcriptome. Mapping so-called splicing quantitative trait loci (sQTLs) can improve the understanding of how genetic variation affects AS and predisposes to IMDs. However, the extent to which sQTLs affect AS in stimulated versus naive macrophages, and how they can affect IMD risk, has not yet been systematically studied.
Here, we map sQTLs in iPSC-derived macrophages obtained from 200 individuals, and exposed to a panel of 12 different stimulants. Gene expression was measured at two timepoints for each condition after stimulation, resulting in 1,953 RNA-seq samples. We identify 5,300 genes that have an sQTL in at least one condition, with a median number of 1,300 such genes per condition. Colocalization analysis of sQTLs and inflammatory bowel disease (IBD) association signals revealed more than 40 loci across all stimulation conditions with posterior probability of sharing a single causal variant > 90%. Most of these high-confidence colocalization events can only be detected in stimulated conditions compared to naive state, and cannot be detected when only eQTL and GWAS signals are colocalized.
Our work highlights the added value of mapping context-aware sQTLs. We show that they provide independent high-confidence colocalizations with IBD risk loci, and that they can enhance the understanding of IMD risk loci as we extend our colocalization analysis to other IMDs.

PrgmNr 1147 - Leveraging single-cell ATAC-seq to identify disease-critical fetal and adult brain cell types

View session detail

Author Block: **S. Kim**[1], K. Jagadeesh[2], K. Dey[2], A. Shen[1], S. Raychaudhuri[3], M. Kellis[1], A. L. Price[2]; [1]Massachusetts Inst. of Technology, Cambridge, MA, [2]Harvard Sch. of Publ. Hlth., Boston, MA, [3]Harvard Med. Sch., Boston, MA

Disclosure Block: **S. Kim:** None.

Identifying disease-critical cell types is a fundamental biological goal (Hekselman et al. 2020 Nat Rev Genet). Single-cell chromatin accessibility (scATAC-seq) and gene expression (scRNA-seq) have characterized cell types at high resolution, and early research on integrating GWAS with scRNA-seq has shown promise, but research on integrating GWAS with scATAC-seq has been limited. In addition, the impact on disease risk of cell types across developmental stages of the brain has not been widely explored.

Here, we integrate GWAS summary statistics from 28 brain-related diseases and traits (avg N=298K) with 3.2 million scATAC-seq and scRNA-seq profiles from 83 cell types spanning fetal and adult brain. We constructed cell-type annotations from scATAC-seq data by annotating SNPs in open chromatin regions, and from scRNA-seq data by annotating SNPs linked to specifically expressed genes using brain-specific enhancer-gene links. We identified disease-critical cell types using stratified LD score regression, assessing statistical significance using FDRIn analyses of fetal brain, we identified disease-critical cell types for 22 traits using scATAC-seq and 8 traits using scRNA-seq, recapitulating known biology. Notable findings using scATAC-seq included significant enrichments of fetal photoreceptor cells for major depressive disorder, fetal ganglion cells for BMI, and fetal astrocytes for ADHD, all of which are supported by known biology but have not previously been identified using genetic data.

In analyses of adult brain, we identified disease-critical cell types for 23 traits using sc-ATAC-seq and 17 traits using scRNA-seq. Notable findings using scATAC-seq included significant enrichments of adult VGLUT2 excitatory neurons for schizophrenia and BDNF excitatory neurons for major depressive disorder. Although fetal brain and adult brain scATAC-seq annotations for the 3 matched cell types were uncorrelated (r=0.00-0.01), the corresponding disease signals were strongly correlated (r=0.52 for -log(P-values)), reflecting partially shared biology. Interestingly, the enrichment of fetal astrocytes for ADHD was not observed in adult astrocytes, consistent with the lower efficiency of adult astrocytes in removing synaptosomes (Sloan et al. 2017 Neuron).

In conclusion, the disease-cell type associations that we identified improve our understanding of the biological mechanisms of complex disease. Cell-type annotations derived from scATAC-seq were particularly powerful in the data that we analyzed, with results only partially shared across fetal vs. adult brain.

PrgmNr 1148 - Computational and functional characterization of the hs737 enhancer in autism

View session detail

Author Block: **E. Padhi**[1], Z. Cheng[2], E. Mehinovic[1], M. Willems[3], L. Pinson[3], S. Redon[4], C. Benech[5], K. Uguen[5], S. Audebert[5], C. Le Marechal[5], C. Ferec[6], S. Efthymiou[7], F. Rahman[8], S. Maqbool[9], R. Maroofian[10], H. Houlden[11], R. Bernier[12], D. U. Gorkin[13,14], B. A. Cohen[15,16], T. Turner[17]; [1]Washington Univ. in St Louis, St Louis, MO, [2]San Diego, CA, [3]HÃ´pital Arnaud de Villeneuve, CHU de MONTPELLIER, MONTPELLIER, France, [4]CHU Brest,Inserm, Univ Brest, Brest, France, [5]CHU Brest, Inserm, Univ Brest, Brest, France, [6]UBO - INSERM - EFS, Brest Cedex 3, France, [7]UCL Inst. of Neurology, London, United Kingdom, [8]Dev. and Behavioral Pediatrics Dept., Inst. of Child Hlth.and Children Hosp., Lahore, Pakistan, [9]Dept. of Neuromuscular Disorders, UCL Inst. of Neurology, London, United Kingdom, [10]london, United Kingdom, [11]Inst Neurology, London, United Kingdom, [12]Univ. of Washington, Seattle, WA, [13]Univ California San Diego, La Jolla, CA, [14]Emory Univ., Atlanta, GA, [15]Washington Univ. in St. Louis, St. Louis, MO, [16]The Edison Family Ctr. for Genome Sci. and Systems Biology, Washington, St. Louis, MO, [17]Washington Univ. Sch. of Med., St. Louis, MO
Disclosure Block: **E. Padhi:** None.

Autism is a complex disorder with a high heritability and currently, based on whole-exome sequencing and array studies, ~10-30% of cases can be explained by variation in protein coding regions. This leaves an appreciable gap in the remaining contribution of genetics to autism. We and others have previously discovered a role for *de novo* variants in noncoding regulatory regions of the genome and predict that they explain 2-5% of individuals with autism. While these noncoding elements (e.g., enhancers) have been implicated in autism and other neurodevelopmental disorders (NDD), there have been few studies identifying specific elements reaching statistical significance and even fewer following up on these elements. Previously, we demonstrated the impact of noncoding variation in hs737 and its link to the critical neurodevelopmental transcription factor, *EBF3*, by analyzing whole-genome sequencing of 10,000 genomes (2,500 families) from the Simons Simplex Collection. Here, we perform deep characterization of hs737 and *EBF3* using computational, functional, and phenotypic analyses. First, to further elucidate the regulatory relationship between hs737 and *EBF3*, we analyzed Hi-C data from human fetal corticogenesis and demonstrate significant interactions between the two elements during neuronal migration. Second, we show by analyzing RNA-seq data from the BrainSpan atlas that *EBF3* expression peaks prenatally and is significantly correlated with many high confidence autism genes. Of these genes, there is a significant enrichment for those involved in chromatin and nucleosome organization, which have been recurrently implicated in NDDs. Finally, we performed phenotypic analyses of 20 probands (7 newly reported here) to demonstrate the consequences of coding and noncoding mutations affecting *EBF3* and find that noncoding mutations are significantly associated with autism while those in *EBF3* are significantly associated with intellectual disability. Currently, we are performing experiments to identify the key regulatory sites within hs737. To do so, we are performing a saturation mutagenesis massively parallel reporter assay for hs737 to gain understand the impact of every possible mutation in hs737. Second, using data from the Vertebrate Genome Project and ENSEMBL, we are performing an evolutionary analysis of hs737 to identify highly conserved nucleotides sites across ~500 species. Comprehensively characterizing noncoding elements involved in autism is an essential next step to understanding its genetics, as doing so will reveal novel regulatory relationships between genes that are not apparent by studying protein-coding regions of the genome.

PrgmNr 1151 - *CAPRIN1* haploinsufficiency causes a novel neurodevelopmental disorder associated with morphological and functional impairment in hiPSCs-derived cortical neurons

View session detail

Author Block: **L. Pavinato**[1,2], A. Delle Vedove[2,3], D. Carli[4], M. Ferrero[1,5], S. Carestiato[1], J. L. Howe[6], E. Agolini[7], D. A. Coviello[8], I. van de Laar[9], P-Y. Billie Au[10], E. Di Gregorio[11], A. Fabbiani[12,13,14], S. Croci[13,14], M. Mencarelli[12], L. P. Bruno[13,14], A. Renieri[12,13,14], R. van Jaarsveld[15,16], A. Childers[17], R. C. Rogers[17], A. Novelli[7], G. B. Ferrero[18], S. W. Scherer[6,19,20], B. Wirth[21], A. Brusco[1,11]; [1]Dept. of Med. Sci., Univ. of Turin, Turin, Italy, [2]Inst. of Human Genetics and Ctr. for Molecular Med. Cologne, Univ. of Cologne, Cologne, Germany, [3]Inst. for Genetics, Univ. of Cologne, Cologne, Germany, [4]Dept. of Publ. Hlth.and Pediatrics, Univ. of Turin, Turin, Italy, [5]Istituto Zooprofilattico Sperimentale del Piemonte, Ligura e Valle d'Aosta, Turin, Italy, [6]The Ctr. for Applied Genomics, Genetics & Genome Biology Program, The Hosp. for Sick Children, Toronto, ON, Canada, [7]Lab. of Med. Genetics, Bambino GesÃ¹ Childrenâ□□s Hosp., Rome, Italy, [8]Lab. of Human Genetics, IRCCS Istituto Giannina Gaslini, Genoa, Italy, [9]Dept. of Clinical Genetics, Erasmus MC, Univ. Med. Ctr. Rotterdam, Rotterdam, Netherlands, [10]Dept. of Med. Genetics, Alberta Children's Hosp. Res. Inst., Univ. of Calgary, Calgary, AB, Canada, [11]Med. Genetics Unit, CittÃ della Salute e della Scienza Univ. Hosp., Turin, Italy, [12]Genetica Medica, Azienda Ospedaliera Univ.ria Senese, Siena, Italy, [13]Med. Genetics, Univ. of Siena, Siena, Italy, [14]Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies, Univ. of Siena, Siena, Italy, [15]Cancer Genomics Netherlands, UMC Utrecht, Utrecht, Netherlands, [16]Molecular Cancer Res., Ctr. for Molecular Med., UMC Utrecht, Utrecht, Netherlands, [17]Greenwood Genetic Ctr., Greenville, SC, [18]Dept. of Clinical and Biological Sci., Univ. of Turin, Turin, Italy, [19]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [20]McLaughlin Ctr., Univ. of Toronto, Toronto, ON, Canada, [21]Inst. of Human Genetics, Ctr. for Molecular Med., Ctr. for Rare Disorders, Univ. of Cologne, Cologne, Germany

Disclosure Block: **L. Pavinato:** None.

Cell cycle Associated PRoteIN 1 (CAPRIN1) is a cytoplasmic ubiquitous phosphoprotein tightly correlated with cellular proliferation. In neurons, CAPRIN1 regulates the transport and translation of mRNAs of proteins involved in synaptic plasticity and interacts with many proteins involved in neurodevelopment, including Fragile-X Mental Retardation Protein (FMRP). Caprin1$^{+/-}$ mice show reduced social interactions and lower response to novelty and, interestingly, *CAPRIN1* has recently emerged as a possible candidate gene for autism spectrum disorders (ASD). We report nine cases carrying loss-of-function (LoF) variants in *CAPRIN1* (one genomic deletion, five nonsense, two invariant-site splicing, one frameshift variant) showing a spectrum of neurodevelopmental phenotypes. All variants were *de novo* except one maternally inherited (maternal phenotype unknown). Overall, language impairment (100%), speech delay (89%), intellectual disability (88%), ADHD (86%), ASD (68%), developmental delay (33%) and seizures (22%) were the main clinical features, together with breathing problems (56%), skeletal malformations (44%), infancy feeding difficulties (33%) and ocular issues (22%). No dysmorphic features were apparent. Expression analysis showed a half dose of the transcript and protein in patients-derived lymphoblasts or fibroblast cells. To further explore the effect of CAPRIN1 haploinsufficiency, we derived cortical neurons from *CAPRIN1*$^{+/-}$ hiPSCs generated by CRISPR/Cas9 mutagenesis. We studied CAPRIN1 temporal expression and localization during neuronal maturation, and found an overall disruption of the neuronal organization, and an increased neuronal death. Global and local alteration of translation in neurons were not observed. Some key proteins of the synaptic plasticity, including FMRP, AMPAR and NMDAR subunits, were down- or upregulated during a time course analysis of neuronal maturation, suggesting a dysregulation of some CAPRIN1-binding partners. Micro-Electrode Arrays (MEAs) measurements of electrical neuronal activity showed lower spike rates and bursts, with an overall

reduced activity. In conclusion, we provide evidence that *CAPRIN1* causes a novel neurodevelopmental disorder with language delay variably associated with ADHD, ASD and intellectual disability. Moreover, we demonstrated that CAPRIN1 half-dose causes morphologic, growth and functional alterations in neuronal cells *in vitro*, supporting haploinsufficiency as pathogenic mechanism.

PrgmNr 1152 - Biallelic *ATG7* variants impair autophagy leading to neurological disease

View session detail

Author Block: **J. Collier**[1], C. Guissart[2], M. OlÃihovÃi[3], S. Sasorith[2], F. Suomi[4], F. Piron-Prunier[5], D. Zhang[6], N. Leboucq[7], A. Bahr[8], S. Azzarello-Burri[8], S. Reich[9], L. SchÃ¶ls[9], T. Polvikoski[3], F. Rivier[7], P. Meyer[7], L. Larrieu[2], H. Alsaif[10], S. Alyamani[10], S. Zuchner[11], I. Barbosa[12], C. Deshpande[13], A. Pyle[3], F. Alkuraya[10], M. Ryten[6], A. Delahodde[5], A. Rauch[8], M. Synofzik[9], R. McFarland[3], T. McWilliams[4], M. Koenig[2], R. Taylor[3]; [1]McGill Univ., Montreal, QC, Canada, [2]Univ. of Montpellier, Montpellier, France, [3]Newcastle Univ., Newcastle, United Kingdom, [4]Univ. of Helsinki, Helsinki, Finland, [5]Inst. de Biologie IntÃ©grative de la Cellule, Gif-sur-Yvette, France, [6]Univ. Coll. London, London, United Kingdom, [7]CHU de Montpellier, Montpellier, France, [8]Univ. of Zurich, Zurich, Switzerland, [9]Univ. of TÃ¼bingen, TÃ¼bingen, Germany, [10]King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia, [11]Univ. of Miami, Miami, FL, [12]King's Coll. London, London, United Kingdom, [13]Guy's and St. Thomas' NHS Fndn. Trust, London, United Kingdom
Disclosure Block: **J. Collier:** None.

Autophagy is an essential developmental and homeostatic process, driving the endolysosomal degradation of protein aggregates, organelles and pathogens. Ablation of autophagy in mouse causes embryonic or perinatal lethality. In humans, dysfunctional autophagy has been implicated in complex diseases including neurodegeneration and cancer, yet congenital autophagy disorders remain exceedingly rare. Using whole exome sequencing we identified pathogenic, biallelic variants in *ATG7*, encoding the principal driver of autophagy, in twelve patients from five families. These patients display complex neurodevelopmental disorders distinguished by selective neurological, neuromuscular and endocrine dysfunction. Notably, cerebellar hypoplasia and a thin posterior corpus callosum were identified in all affected patients who were assessed by neuroimaging. Contrasting conditional *Atg7* deletion in mouse which causes perinatal lethality, patients survive into adult life despite loss of ATG7 protein. Fibroblasts from affected individuals from each family displayed diminished autophagy, and expression of mutated *ATG7* failed to rescue autophagy-deficient model systems, supporting the pathogenicity of these variants. ATG7-deficient patient muscle revealed myopathic and inflammatory changes, together with subsarcolemmal p62/SQSTM1 accumulation. Despite loss of ATG7, autophagic structures were readily detected in patient fibroblasts and muscle, suggesting that ATG7-independent autophagosome biogenesis pathways may support basal autophagic degradation in human cells. Our study provides the first clinical, genetic and mechanistic demonstration that mutated *ATG7* leads to neurodevelopmental disease in humans, who can survive for decades with defective canonical autophagy. Importantly, two patients with undetectable ATG7 protein display a relatively mild phenotype, revealing that human life is compatible with the absence of a nonredundant core autophagy gene, thereby challenging current perceptions regarding the relationship between autophagy and human health and disease.

PrgmNr 1153 - ÂÂBi-allelic variants in the autophagy gene *ATG4D* are associated with a pediatric neurological disorder characterized by hypotonia, joint laxity, and delayed speech and motor development

View session detail

Author Block: **M. Morimoto**[1,2], M. Davids[1,2], P. M. Zerfas[3], V. Bhambhani[4], A. Lehman[5], J. L. Murphy[1], E. Macnamara[1], C. Wang[1], Undiagnosed Diseases Network, S. Reichert[6], A. Thurm[7], W. J. Introne[2], W. A. Gahl[1,2], C. J. Tifft[1,2,8], M. C. V. Malicdan[1,2]; [1]NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, NIH, Bethesda, MD, [2]Med. Genetics Branch, Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [3]Diagnostic and Res. Services Branch, Office of Res. Services, NIH, Bethesda, MD, [4]Dept. of Med. Genetics, Childrenâ□□s Hosp. and Clinics of Minnesota, Minneapolis, MN, [5]Dept. of Med. Genetics, Faculty of Med., Univ. of British Columbia, Vancouver, BC, Canada, [6]Dept. of Clinical Genetics, Virginia Commonwealth Univ. Hlth., Richmond, VA, [7]Neurodevelopmental and Behavioral Phenotyping Service, Office of the Clinical Director, Natl. Inst. of Mental Hlth., NIH, Bethesda, MD, [8]Office of the Clinical Director, Natl. Human Genome Res. Inst., NIH, Bethesda, MD

Disclosure Block: **M. Morimoto:** None.

Introduction: Macroautophagy is a dynamic and highly conserved process that regulates the degradation and recycling of cellular components. Autophagosomes are double-membraned organelles that deliver cellular components to the lysosomes for degradation. Defective autophagy has been shown to contribute to the pathogenesis of various human diseases. We describe three individuals from two unrelated families with a pediatric neurological disorder with bi-allelic variants in *ATG4D* that encodes one of the four ATG4 isoforms that process the LC3 and GABARAP protein subfamilies required for autophagosome biogenesis. Notably, a homozygous *ATG4D* missense variant is associated with a canine neurodegenerative disease. **Methods:** A detailed clinical evaluation was performed on each affected individual. Whole exome sequencing and family-based genomics were performed to identify candidate variants; variants were prioritized based on their frequency in population databases, level of conservation, and predicted deleteriousness. Expression analyses and functional assays to assess autophagosome biogenesis and autophagic flux were performed on patient cells and ATG4D-deficient HeLa cells generated by CRISPR-Cas9. Further, an *in vitro* assay using purified recombinant proteins was performed to evaluate the processing efficiency of each individual ATG4D variant on a known target of ATG4D, GABARAPL1. **Results:** Three individuals presented with a pediatric neurological disorder characterized by hypotonia, joint laxity, and delayed speech and motor development. Rare, conserved, and likely deleterious bi-allelic missense or frameshift variants were identified in *ATG4D* segregating with the disease. *ATG4D* mRNA expression and ATG4D protein levels in patient cells were comparable to controls. Autophagosome biogenesis and autophagic flux in patient cells and ATG4D-deficient HeLa cells were similar to controls, suggesting functional redundancy due to the other ATG4 isoforms. In order to functionally assess ATG4D activity in the absence of the other ATG4 isoforms, an *in vitro* priming assay was performed and demonstrated diminished GABARAPL1 processing efficiency for three of the four *ATG4D* variants. **Conclusion:** The clinical, bioinformatic, and functional data, together with the previously described role of ATG4D in autophagy and its association with neurodegenerative disease in other organisms, provide evidence that bi-allelic loss-of-function variants in *ATG4D* likely underlie the pathogenesis of this novel neurological disorder.

PrgmNr 1154 - Large-scale, multi-ethnic resource of gene, isoform, and splicing regulation in the developing human brain

View session detail

Author Block: **C. Wen**[1], R. Dai[2], P. F. Przytycki[3], M. Kim[4], A. Bhattacharya[1], P. Zhang[1], R. L. Walker[5], PsychENCODE Consortium, D. Pinto[6], K. S. Pollard[3], C. Liu[2], M. J. Gandal[1]; [1]UCLA, Los Angeles, CA, [2]SUNY Upstate Med. Univ., Syracuse, NY, [3]Univ California San Francisco, San Francisco, CA, [4]David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [5]Broad Inst., Boston, MA, [6]Icahn Sch. of Med. at Mount Sinai, New York, NY

Disclosure Block: **C. Wen:** None.

Genome-wide associations (GWAS) for neuropsychiatric disorders including schizophrenia (SZ) and autism show strong enrichment among fetal brain regulatory elements. This has prompted several recent efforts to systematically characterize the impact of genetic variation on gene regulation in the developing human brain through expression, isoform, and splicing quantitative trait loci (eQTL, isoQTL, sQTL) mapping. However, to date, the scale of individual fetal brain studies has been relatively small, and lack of consistent processing has limited the ability to compare results across studies or with larger adult brain xQTL datasets. Moreover, there has been no systematic comparison of xQTL results across trimesters or ancestries.

Here we uniformly processed and mega-analyzed 5 fetal brain genomic datasets, 682 subjects in total, across European (N=292), African-American (N=164), and Admixed American ancestries (N=145), spanning all three developmental trimesters. Gene and isoform expression levels were quantified from raw RNAseq reads using Salmon and splicing levels were quantified using Leafcutter. SNP genotypes were imputed into the TOPMed reference panel. xQTL was comprehensively analyzed via FastQTL with adaptive permutations and multiple testing corrections, followed by fine mapping with SuSiE. Cell type-deconvolution with expression-based Swcam and chromatin accessibility-based CellWalker enabled the detection of cell type-specific and interacting genetic regulation. We identified 10094 genes, 22891 isoforms, and 37706 introns--many specific to fetal brain--showing significant cis-regulation. Compared with adult brain datasets from GTEx and PsychENCODE, we identified 5487 and 6325 new eGenes, respectively. cis-eQTL effect sizes are highly concordant between ancestries, with all pair-wise Spearman Ï greater than 0.83. We found substantial enrichment of psychiatric GWAS signals among all fetal brain xQTL annotations. Integration of fetal brain xQTL with GWAS via transcriptome-wide association study (TWAS) prioritized 96 candidate risk genes for SZ, 34 more than the previously largest fetal brain dataset.

By harmonizing and analyzing fetal brain datasets, we characterized xQTL and discovered that fetal brain regulatory elements are associated with neuropsychiatric disorders and other brain-related traits.

PrgmNr 1155 - Biallelic variants in *CELSR3* cause a syndrome comprising CAKUT, anorectal and CNS anomalies

View session detail

Author Block: J. Stegmann[1,2], J. C. Kalanithy[1,2], G. C. Dworschak[1,2,3], E. Mingardo[1,2], G. Hagelueken[4], M. Geyer[4], S. Seltzsam[5], F. Boschann[6], J. A. Rosenfeld[7,8], E. Mizerik[7,9], K. Myers[10], M. Morleo[11,12], R. J. Olson[13], J-H. Chae[14], A. R. Janecke[15,16], I. Vogel[17,18], G. N. Wilson[19], J. L. Cohen[20], L. E. Walsh[21], A. Lai[22], M. Begemann[23], J. E. Posey[7], J. R. Lupski[7,24,9,25], B. Odermatt[2,26], H. Reutter[1,27]; [1]Inst. of Human Genetics, Med. Faculty, Univ. of Bonn, Bonn, Germany, [2]Inst. of Anatomy and Cell Biology, Med. Faculty, Univ. of Bonn, Bonn, Germany, [3]Dept. of Pediatrics, Univ. Hosp. Bonn, Bonn, Germany, [4]Inst. of Structural Biology, Univ. Hosp. Bonn, Univ. of Bonn, Bonn, Germany, [5]Boston Childrenâ□□s Hosp. and Harvard Med. Sch., Boston, MA, [6]Inst. of Med. Genetics and Human Genetics, CharitÃ© Univ.smedizin Berlin, corporate member of Freie Univ. Berlin and Humboldt-Univ. zu Berlin, Berlin, Germany, [7]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [8]Baylor Genetics Lab., Houston, TX, [9]Texas Children's Hosp., Houston, TX, [10]Ctr. for Cardiovascular Res., Nationwide Childrenâ□□s Hosp., Dept. of Pediatrics, Ohio State Univ., Columbus, OH, [11]Telethon Inst. of Genetics and Med. (TIGEM), Pozzuoli, Naples, Italy, [12]Telethon Undiagnosed Diseases Program, Naples, Italy, [13]Ctr. for Individualized Med., Mayo Clinic, Rochester, MN, [14]Dept. of Genomic Med., Dept. of Pediatrics, Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, [15]Dept. of Pediatrics, Med. Univ. of Innsbruck, Innsbruck, Austria, [16]Div. of Human Genetics, Med. Univ. of Innsbruck, Innsbruck, Austria, [17]Ctr. for Fetal Diagnostics, Dept. of Clinical Med., Aarhus Univ., Aarhus, Denmark, [18]Dept. of Clinical genetics, Aarhus Univ. Hosp., Aarhus, Denmark, [19]Dept. of Pediatrics, Texas Tech Univ. Hlth.Sci. Ctr., Lubbock and KinderGenome Med. genetics practice, Dallas, TX, [20]Div. of Med. Genetics, Dept. of Pediatrics, Duke Univ., Durham, NC, [21]Pediatric Neurology, Riley Hosp. for Children Indiana Univ. Hlth., Indianapolis, IN, [22]Div. of Genetics and Genomics, Dept. of Pediatrics, Boston Childrenâ□□s Hosp., Boston, MA, [23]Inst. of Human Genetics, Med. Faculty, RWTH Aachen Univ., Aachen, Germany, [24]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [25]Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, [26]Inst. of Neuroanatomy, Med. Faculty, Univ. of Bonn, Bonn, Germany, [27]Dept. of Neonatology and Pediatric Intensive Care, Univ. Hosp. Bonn, Bonn, Germany

Disclosure Block: J. Stegmann: None.

CELSR3 encodes the adhesion G protein-coupled receptor (aGPCR) ADGRC3 of the cadherin superfamily. It plays an important role in neuronal migration and planar cell polarity signaling. All known Cadherin EGF LAG seven-pass G-type receptors (CELSR1-3) are predicted to have various functions in epithelial cells and in the central nervous system (CNS). *Celsr1-3* mRNA patterns in mouse embryo revealed distinct expression in the developing CNS and during ureteric budding**.** Here we describe 14 families with congenital anomalies of the kidney and urinary tract (CAKUT), anorectal and CNS anomalies. These phenotypes either occurred in combination with each other or isolated among the affected. Assembly of these families was facilitated by GeneMatcher. The spectrum of kidney anomalies among the patients comprised anatomical and functional aberrations. In all patients, biallelic variants in *CELSR3* were found. Protein modelling of selected missense variants implicated the disruption of the normal protein structure, hinting towards a malfunction of the altered protein. To examine the function of ADGRC3 during early development and to gain insights into the embryonic architecture of aberrant kidney formations, we established a zebrafish larvae (zfl) knockdown model. Transient knockdown of *celsr3* in fluorescent reporter lines specific to the neuronal and urinary systems in zfl showed a disturbed embryonic development of the urinary tract and the CNS. Our molecular studies in humans and functional analysis of Celsr3 in zfl provide insights into *CELSR3* as a potential human disease gene for a novel syndrome comprising CAKUT, anorectal and CNS anomalies of variable expression. Grant: J.D.S.: BonnNI grant Q614.2454. G.C.D.: BONFOR grant

O-120.0001 and Herbert-Reeck foundation (2019). HR: German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) RE 1723/5-1. J.R.L.: Jointly funded National Human Genome Research Institute (NHGRI) and National Heart, Lung, and Blood Institute (NHLBI) grant to the Baylor-Hopkins Center for Mendelian Genomics [UM1 HG006542].

PrgmNr 1156 - Statistical framework uncovers deep intronic splice gain variants implicated in undiagnosed cases

View session detail

Author Block: **S. Kobren**[1], R. M. Reimers[2], D. Barnum[2,1], J. Krier[2], I. Kohane[1], R. Sherwood[3], S. Sunyaev[4]; [1]Harvard Med. Sch., Boston, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]Brigham and Women's Hosp., Harvard Med Sch., Boston, MA, [4]Harvard Med. Sch. and Brigham & Women's Hosp., Boston, MA

Disclosure Block: **S. Kobren:** None.

Non-coding splice-altering variants are thought to be implicated in over 15% of rare genetic disorders and are hypothesized to be particularly relevant in cases where targeted or whole exome sequencing failed to reveal disease causative variants. However, identifying deeply intronic variants that create pathogenic, cryptic splice site gains still represents a formidable challenge, especially in cases where obtaining RNA-seq data from relevant tissue types (e.g., brain) is impossible. Existing methods to predict the splice-altering impact of DNA variants directly from whole genome sequencing are plagued by a high rate of false positives as well as a nontrivial rate of false negatives, reducing their immediate utility in diagnostic workflows. Here, we develop a precise statistical model that incorporates intronic SpliceAI scores, coding region pathogenicity predictions, as well as likelihoods of specific, rare variants based on parental variant occurrence and an underlying mutational model in order to assess the significance of compound heterozygous variant pairs observed in recessive cases. When applied to a harmonized dataset of over 800 patients with suspected Mendelian conditions from the Undiagnosed Diseases Network, our approach recapitulates known diagnoses with high precision as well as significantly ranks previously uninvestigated splice-altering variants that perturb clinically relevant genes in undiagnosed cases. Our process includes a systematic clinical evaluation of top candidates, a high-throughput massively parallel splicing assay (MPSA) to verify the impact of intronic variants across a wide range of SpliceAI scores, as well as RNA-guided base alterations in cell line models to measure phenotypes of potentially diagnostic variants. Our results demonstrate that our statistical framework can accurately and rapidly detect clinically viable and verifiable cryptic splice site gains with etiological roles across cohorts of chronically undiagnosed cases.

PrgmNr 1159 - The impact of copy number variants on complex human traits

View session detail

Author Block: **C. Auwerx**[1,2,3,4], M. Lepamets[5,6], M. Sadler[2,3,4], E. Porcu[1,2,3,4], R. MÃ¤gi[6], A. Reymond[1], Z. Kutalik[2,3,4]; [1]Ctr. for Integrative Genomics, Univ. of Lausanne, Lausanne, Switzerland, [2]Dept. of Computational Biology, Univ. of Lausanne, Lausanne, Switzerland, [3]Swiss Inst. of Bioinformatics, Lausanne, Switzerland, [4]Univ. Ctr. for Primary Care and Publ. Hlth., Lausanne, Switzerland, [5]Inst. of Molecular and Cell Biology, Univ. of Tartu, Tartu, Estonia, [6]Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia

Disclosure Block: **C. Auwerx:** None.

Copy number variations (CNVs) are potent phenotypic modifiers associated with several rare genomic syndromes but their impact on complex traits remains understudied. To fill this gap, we called CNVs in 331â□□522 white, unrelated UK Biobank participants and performed genome-wide association studies (GWASs) between the copy-number of CNV-proxy probes and 57 continuous traits, revealing 131 independent signals across 47 phenotypes. Besides recapitulating well-known associations (e.g. 16p11.2 and weight), our analysis revealed the pleiotropic impact of recurrent CNVs, with 26 and 16 traits associated with 16p11.2-BP4-BP5 and 22q11.21, respectively. Forty CNV signals (31%) overlapped with previously identified GWAS signals for the same trait. For instance, deletion of urate transporter scaffold protein-encoding *PDZK1* decreased urate levels ($\hat{I}^2_{del}$ = -48.3 Î¼mol/L, p = 5.8e-13), with a 2.6-fold stronger effect in males (p_{diff} = 2.1e-3). Many associations mapped to rare disease regions, suggesting variable expressivity and a broad impact of these loci in the general population. For example, heterozygous deletion of the region encoding the hepatic transporters OATP1B1/3 associated with increased total bilirubin ($\hat{I}^2_{del}$ = 3.1 Î¼mol/L, p = 2.2e-13), whereas homozygous carriers affected with Rotor syndrome present with severe hyperbilirubinemia. Our approach also highlighted new gene functionalities: The copy-number of the region encompassing *MARF1*, a gene essential to murine oogenesis, correlated negatively with age at menarche ($\hat{I}^2_{mirror}$ = -0.6 years, p = 8.5e-15) and menopause ($\hat{I}^2_{dup}$ = -1.8 years, p = 1.7e-6), suggesting a conserved role in female reproductive timing. Moving beyond single CNVs, we found that an individualâ□□s CNV burden negatively impacted 35 of the 57 assessed traits, leading among others to increased adiposity and liver/kidney damage biomarkers, and decreased intelligence and physical capacity. Thirty traits remained burden-associated after correcting for the impact of significantly associated CNV regions, pointing to a polygenic CNV-architecture underlying complex traits and suggesting the presence of additional associations, which we currently lack the power to detect. Finally, we showed that an individualâ□□s CNV burden was negatively associated with both parental lifespan ($\hat{I}^2_{burden}$ = -0.18 years/Mb, p = 1.1e-7) and age (proxy for survivorship; $\hat{I}^2_{burden}$ = -0.21 years/Mb, p = 1.4e-5), suggesting that the deleterious impact of CNVs contributes to decreased longevity. Together, the numerous identified associations suggest that CNVs play an important role in shaping complex traits and that their study can reveal new biological insights.

PrgmNr 1160 - PAV: An assembly-based approach for discovering structural variants, indels, and point mutations in long-read phased genomes

View session detail

Author Block: **P. Audano**[1], P. Ebert[2], W. Harvey[3], Q. Zhu[1], Y. Mao[3], The Human Genome Structural Variant Consortium, C. Lee[1], T. Marschall[2], C. R. Beck[1,4,5], E. E. Eichler[3]; [1]The Jackson Lab. Inst. for Genomic Med., Farmington, CT, [2]Heinrich Heine Inst., DÃ¼sseldorf, Germany, [3]Univ. of Washington, Seattle, WA, [4]Dept. of Genetics and Genome Sci., Univ. of Connecticut Hlth.Ctr., Farmington, CT, [5]Inst. for Systems Genomics, Univ. of Connecticut Hlth.Ctr., Farmington, CT
Disclosure Block: **P. Audano:** None.
While variant discovery has been dominated by read alignments, phased genome assemblies offer greater contiguity across large repeats and structural variants (SVs) making them ideal for identifying a wider range of variant sizes over a larger proportion of the genome. Recent advances have increased the fidelity and contiguity of long-read assemblies, improved methods are phasing them into distinct haplotypes, and new tools are required to leverage these advances for characterizing human genetic variation more completely. We developed the phased assembly variant (PAV) pipeline designed for fast and accurate discovery of SVs, indels, and point mutations using assembled contigs from human genomes instead of read-based discovery. To control for false calls and to refine breakpoints for large SVs, we have introduced an alignment trimming step which eliminates 11.5 Mbp of redundantly mapped bases per assembled haplotype on average. By implementing a novel inversion detection algorithm that does not rely exclusively on split alignments, we can now resolve small inversions that are difficult to detect with current approaches while eliminating erroneous SVs and dense clusters of false indels and SNVs over these inverted sites. Per sample, we now routinely discover more than 24,000 structural variants per human genome including 55 inversions between 300 bp and 80 kbp more than doubling inversion calls from previous assembly-based methods. PAV was the foundation of a recently published callset constructed from 35 phased samples (70 haplotypes) using Strand-seq phased HiFi and CLR long-read assemblies. We are currently applying PAV to additional human genomes from the Human Pan Genome Reference Consortium as well as mouse and nonhuman primate genomes producing the most comprehensive set of sequence-resolved structural variants to date. Due to the reliability of the alignment and trimming steps, PAV accurately places SV junctions, which can be used to probe for mechanistic signatures such as breakpoint homology. PAV (v.1.1) has been further optimized to improve characterization of larger events and to more accurately delineate breakpoints to better understand mechanisms of origin. When paired with improved assembly methods, PAV has further increased sensitivity for inversions, and is now detecting up to 86 events per sample with new calls concentrated in larger sizes up to hundreds of kbp. The results of this work will serve as the basis for a new database of human structural variation critical for the discovery of pathogenic genetic variants as long-read sequence datasets begin to be more routinely applied to unsolved disease samples.

PrgmNr 1161 - Classes of rare and common variants differentially contribute to variably expressive phenotypes in complex disorders

View session detail

Author Block: **M. C. Jensen**[1], A. Tyryshkina[1], L. Pizzo[1], C. Smolen[1], V. Pounraja[1], E. Huber[1], C. Taylor[2], V. M. Rajagopal[3], H. Song[1], 16p12.1 Deletion Consortium, S. Girirajan[1]; [1]Pennsylvania State Univ., University Park, PA, [2]Geisinger Hlth.System, Lewisburg, PA, [3]Aarhus Univ, Aarhus, Denmark
Disclosure Block: **M.C. Jensen:** None.
Copy-number variants associated with complex disorders, such as autism and intellectual disability, are often characterized by extensive phenotypic variability. For example, the 520-kbp deletion on chromosome 16p12.1 contributes to schizophrenia risk and decreased intelligence among adults in the general population, and is also enriched among children with severe developmental delay, autism, and congenital features. As >90% of severely affected children inherit the deletion from a mildly affected parent, we hypothesized that the deletion sensitizes the genome for neuropsychiatric traits, while variants elsewhere in the genome, or â□□second-hitsâ□□, determine the ultimate phenotypic outcome. We therefore developed a framework to systematically quantify the roles of rare and common â□□second-hitâ□□ variants towards variable phenotypes of complex disorders, and used this framework to perform whole-genome sequencing and deep clinical phenotyping on 410 individuals in 131 families with the 16p12.1 deletion. We found that carrier children were enriched for rare variants, including short tandem repeats (STRs, p=0.047), loss-of-function (LOF) variants (p=0.041), and missense SNVs in genes intolerant to variation (p=0.030), compared to their carrier parents. Interestingly, carrier parents had higher polygenic risk for schizophrenia compared to their carrier children ($p=2.98Ã□10^{-3}$), suggesting differential contributions for rare and common variants towards early-onset and late-onset clinical features. Using mixed logistic models, we next assessed the effects of >25 rare variant classes towards six distinct developmental phenotypic domains in carrier children. For example, rare deletions (OR=15.45, p=0.011) and splice-site variants (OR=2.64, p=0.037), showed significant effects towards intellectual disability, while STRs were enriched in children with nervous system defects ($OR=3.20, p=8.73Ã□10^{-3}$) and LOF variants were enriched in children with congenital malformations (OR=2.01, p=0.014). We also observed correlations between rare deleterious coding variants and IQ (R=-0.326) and autism severity scores (R=0.283) among the deletion carriers. Finally, we identified enrichments for combinations of genes with â□□second-hitsâ□□ among carrier children with specific phenotypes, such as *ARHGEF19* and *ASPM* for psychiatric traits and *MYO7A* and *HYDIN* for musculoskeletal defects. Overall, we identified specific patterns of rare and common variants for distinct phenotypic domains among 16p12.1 deletion carriers, highlighting the utility of dissecting the genetic etiology of variable phenotypes to better understand complex disorders.

PrgmNr 1162 - Modeling haplotype sharing powers CNV detection and reveals large-effect phenotype associations

View session detail

Author Block: **M. Hujoel**[1], M. A. Sherman[2], A. Barton[3], P-R. Loh[1]; [1]Brigham and Women's Hosp. / Harvard Med. Sch., Boston, MA, [2]Massachusetts Inst. of Technology, Marblehead, MA, [3]Harvard Med. Sch., Boston, MA

Disclosure Block: **M. Hujoel:** None.

The effects of most copy number variants (CNVs) on human phenotypes are unknown. Existing methods for detecting CNVs from biobank-scale SNP-array data can only sensitively detect large CNVs (~50kb+). To address this challenge, we developed a new approach leveraging the idea that inherited, population-polymorphic CNVs are usually carried by multiple individuals within a large cohort, such that combining SNP-array data across multiple potential carriers of a CNV can greatly increase detection power.

-

Applying this approach to 452,520 UK Biobank participants detected an average of 31 CNVs per individual (~18 deletions and ~13 duplications), an order of magnitude more than previous analyses. Detected duplications were on average longer than deletions (84.7kb vs. 26.3kb), such that duplications tended to affect more total base pairs than deletions (900kb vs. 430kb). Validation using whole-genome sequencing of 48 participants indicated 80% sensitivity for detecting low-frequency CNVs >5kb that spanned at least 2 SNP-probes, with a false discovery rate -

Association and fine-mapping analyses of these CNVs with 58 quantitative traits identified 301 independent associations (*P* -8) - involving CNVs at 101 loci - that could not be explained by linkage disequilibrium with any nearby SNP. Most associated CNVs were very rare (MAFTFRC, which encodes the transferrin receptor, associated with a 2.24 (s.e. 0.22) SD decrease in mean corpuscular hemoglobin (P=3 x 10^{-22}) and increased anemia risk (OR = 5.2 (1.1-23.7)), and a very rare deletion of *DIS3L2* exon 9 (MAF=0.03%) previously implicated in autosomal recessive Perlman syndrome associated with a 1.1 inch (s.e. 0.1 inch) decrease in height in heterozygous carriers (P=4 x 10^{-22}). Rare CNVs also created extended allelic series at *JAK2* and *HBA* including deletions or duplications of distal enhancers that associated with much stronger phenotypic effects (>0.5 SD) than SNPs within these regulatory elements. Deletion and duplication burdens of non-syndromic CNVs each associated with decreased height and years of education (*P*

PrgmNr 1163 - Multi-tool discovery and joint genotyping of structural variations in 138,134 multi-ethnic TOPMed WGS samples

View session detail

Author Block: **G. Jun**[1], F. J. Sedlazeck[2], G. A. Metcalf[2], A. C. English[2], V. Menon[2], T. Blackwell[3], H. M. Kang[3], A. V. Smith[3], R. A. Gibbs[2], G. R. Abecasis[3], E. Boerwinkle[1]; [1]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Baylor Coll. of Med., Houston, TX, [3]Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **G. Jun:** None.
Large-scale whole-genome sequencing (WGS) studies enable the evaluation of the functional impact of structural variations (SVs) on multiple human phenotypes and conditions. SVs are 50bp or larger genomic alterations and have been shown to be the main driver of genomic diversity with clear roles in human diseases or other phenotypes. There remains substantial technical challenges in discovery and genotyping of SVs from short-read sequence data, especially at-scale. A plethora of SV detection methods have been developed over the last decade and these methods have strengths identifying different SV types and sizes. We developed Parliament2, a SV discovery and merging pipeline that harmonizes calls from multiple SV detection software into a single SV callset that shows better specificity and sensitivity than any single method. We also developed an efficient joint-genotyping method, muCNV, that reduces false discoveries that accumulate across methods and large sample sizes. The Parliament2-muCNV pipeline has been employed to generate a comprehensive catalog of SVs for genetic association analyses across 138,134 multi-ethnic TOPMed WGS samples. We identified a total of 466,800 SVs, including 231,817 deletions, 197,412 duplications/CNVs, and 37,571 inversions. As expected, the majority of SVs were rare, with almost 46% being singletons. On average, an individual carries 3,303 deletions, 3,570 duplications, and 185 inversions. To estimate genotyping accuracy, we evaluated non-reference Mendelian inconsistency rates using 11,580 trios. The estimated error rates were 0.29% for deletions, 0.83% for biallelic duplications, and 3.1% for inversions. *De novo*heterozygote rates also showed similar numbers: 0.45%, 0.66%, and 4.2% for deletions, duplications, and inversions, respectively, demonstrating that the callset is suitable for genetic association analyses. The TOPMed program is ongoing and growing in sample size, and we expect an even larger and highly accurate SV callset at the time of presentation. These SV data are available for download from dbGaP to researchers with an approved TOPMed manuscript proposal and currently being analyzed for associations with a number of different traits. Assessing contributions of SVs to common and complex human traits and risk factors will further advance our understanding of genetics and biology of heart, lung, blood, and sleep disorders, especially when combined with transcriptomic, epigenetic, and metabolomic resources from the TOPMed program.

PrgmNr 1164 - Sequencing individual genomes with recurrent deletions reveals allelic architecture and disease loci for autosomal recessive traits

View session detail

Author Block: **P. Liu**[1,2], B. Yuan[3], K. V. Schulze[4], N. Assia Batzir[5], J. Sinson[1], H. DAI[6], W. Zhu[1], F. Bocanegra[7], C-t. Fong[8], J. Holder[9], J. Nguyen[10], C. P. Schaaf[11], Y. Yang[12], W. Bi[13], C. M. Eng[14], C. Shaw[14], J. R. Lupski[15]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics, Houston, TX, [3]Seattle Children's Hosp., Seattle, WA, [4]Houston, TX, [5]Baylor Coll. Med., Tel-Aviv, Israel, [6]Baylor Coll. Med., Houston, TX, [7]Inst. de Referencia Andino, BogotÃi, Colombia, [8]Univ of Rochester, Rochester, NY, [910]Dallas, TX, [11]Univ. of Cologne, KÃ¶ln, Germany, [12]AiLife Diagnostics, Pearland, TX, [13]Pearland, TX, [14]Baylor Col Med., Houston, TX, [15]Baylor Col Med, Houston, TX
Disclosure Block: **P. Liu:** Other; Baylor Genetics.

In medical genetics, discovery and characterization of new â□□disease genesâ□□ and disease associated rare variant alleles depends on genetic reasoning, study designs, and subsequently patient ascertainment that can reveal these novel findings and parsimoniously explain potential variant/gene contribution to the disease trait. Here, we present new insights to enhance discovery in the challenging context of recessive disease traits and bi-allelic disease. We demonstrate through computational analyses that a collection of 30 large recurrent genomic deletions scattered throughout the human genome likely contribute to more than 20% of disease load for over 2% of all known â□□recessive disease genesâ□□. Moreover, our modeling suggests that these deletion alleles are responsible for the majority of recessive disease burden in at least 70% of the segmental regions triggered by the recurrent rearrangement NAHR mechanism; which occurs *de novo* at such loci at a relatively high mutation rate (rates of 10^{-5} to 10^{-6}). We hypothesize that genetic sequencing of the hemizygous chromosomal regions in *trans* to these large deletions is an under-utilized approach for allele and gene discovery as well as an opportunity to improve clinical molecular diagnosis. To explore our hypothesis, we performed meta-analyses for all literature-reported affected patients with the 23 recessive disease genes whose carrier burden are predicted to be overwhelmingly from large recurrent genomic deletions. Our data and analyses support the contention that current diagnostic efforts for personal genomes with large recurrent deletions do not adequately exploit genome sequencing for its potential role in discovery and molecular diagnosis. Moreover, by reanalysis of previously undiagnostic exome sequencing (ES) data on 69 subjects harboring 26 distinct recurrent deletions, probable new diagnostic variants were uncovered in genes including *COX10*, *ERCC6, PRRT2* and *OTUD7A,* thereby demonstrating the potential clinical and discovery impact of our proposed approach. Finally, we suggest that analyses of population specific allele spectra may provide important insights for study design. Findings from this study support the contention that more whole genome sequencing (WGS) may further resolve molecular diagnoses and provide evidence for multi-locus pathogenic variation (MPV). Such analyses benefit all stakeholders in both research and patient clinical care.

PrgmNr 1167 - Web-based analysis and collaboration tool, seqr, for rare disease genomics

View session detail

Author Block: **L. Pais**[1], H. Snow[1], S. Zhang[1], S. DiTroia[1], E. O'Heir[1], C. Austin-Tse[1,2], B. Weisburd[1], M. Wilson[1], S. Baxter[1], D. MacArthur[3], A. O'Donnell-Luria[1,4], H. Rehm[1,2]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Garvan Inst. of Med. Res. and Murdoch Children's Res. Inst., Sydney, Australia, [4]Boston Children's Hosp., Boston, MA
Disclosure Block: **L. Pais:** None.
The accelerated adoption and decreased costs of genomic sequencing have exponentially increased the amount of data available for rare disease analyses. To identify causal variants in these large datasets, powerful filtering and decision support tools are needed that can be easily used by researchers. While several tools for rare disease analysis exist, few are freely available. We have addressed this need through seqr - an open source, web-based tool for family-based monogenic disease analysis that allows researchers to work collaboratively to search and annotate genomic callsets. To enable more widespread access to this tool, seqr is also now available for any joint-called VCF files loaded on the AnVIL platform (anvilproject.org). To date, seqr has enabled analyses of over 10,000 unsolved cases in collaboration with the Broad Institute Center for Mendelian Genomics and facilitated the diagnosis of more than 3,800 cases and discovery of over 300 novel disease genes across our collaborative network of 45 research teams from 57 countries.
The user-friendly seqr interface allows researchers to filter a richly annotated variant call file using pre-defined default searches or their own customized search parameters. The platform supports efficient interrogation of candidate single nucleotide variants, small insertions and deletions, and copy number variants through integration of interactive read visualization, variant and gene-level annotations from multiple internal and external resources, and support for downstream data sharing. We will present examples of cases solved in seqr involving phenotype expansions, incomplete penetrance, genomic hotspots and deep intronic variants. We will also describe the seqr matchmaker interface that allows submission of candidates and metadata to Matchmaker Exchange for gene discovery, including tracking over 6,500 match communications to date. Finally, we will show how seqr can be used for project management and data sharing purposes. This presentation will demonstrate seqrâ□□s advanced capabilities and upcoming features for genomic analyses, and how researchers can utilize it in their respective rare disease research pipelines.

PrgmNr 1168 - Mendelian diseases and their role in a pediatric neurodegeneration cohort

View session detail

Author Block: **B. DiSanza**[1], R. Guo[1], F. D. Mentch[2], E. E. Will[1], E. J. Bhoj[1], R. C. Ahrens-Nicklas[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Childrens Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **B. DiSanza:** None.
Pediatric neurodegenerative diseases are an uncommon and heterogeneous category of disorders that affect approximately 0.6 out of every 1,000 live births. Here, we further clarify the Mendelian etiology of neurodegenerative diseases observed in an unbiased cohort of pediatric patients, as well as the clinical implications of receiving the appropriate genetic diagnosis. Patients enrolled in the Center for Applied Genomics (CAG) Biobank at the Children's Hospital of Philadelphia were filtered through an algorithm to generate a cohort of pediatric patients with neurodegenerative phenotypes. We discovered 76 cases with a neurodegenerative phenotype out of ~100,000 pediatric patients enrolled. 7 patients were excluded after manual chart review due to secondary neurologic symptoms caused by other medical conditions or a total absence of neurologic symptoms. 42 patients (60.9%) had a genetic diagnosis, whereas 27 were undiagnosed (39.1%). We identified 32 distinct illnesses among those diagnosed. Rett syndrome, mitochondrial illnesses (Leigh syndrome), neuronal ceroid lipofuscinoses, X-linked adrenoleukodystrophy, and Aicardi Goutieres syndrome were among the most common monogenic diagnoses. In 39/42 (92.8%) of patients, obtaining a genetic diagnosis resulted in a change in clinical care, including the monitoring of known related problems, drug management, and/or enrollment in a clinical trial. We then performed genomic and metabolomic analysis to integrate in the undiagnosed cohort of pediatric neurodegeneration patients. These neurodegeneration cases differed significantly from healthy controls in several metabolites identified through untargeted metabolomics. In summary, the disorders uncovered in this cohort indicate the wide range of pathophysiology and diseases contributing to pediatric neurodegeneration. However, 39.1% of pediatric patients remained undiagnosed despite genetic testing for neurologic symptoms, thus emphasizing the importance of ongoing research in this field and the need for further studies. Techniques involving integrated genomic, transcriptomic, and metabolomic analysis should be utilized in gene discovery.

PrgmNr 1169 - A genome-first approach to rare variants in hypertrophic cardiomyopathy genes *MYBPC3* and *MYH7* in a medical biobank

View session detail

Author Block: **J. Park**, E. A. Packard, M. G. Levin, R. L. Judy, Regeneron Genetics Center, S. M. Damrauer, S. M. Day, M. D. Ritchie, D. J. Rader; Perelman Sch. of Med. at the Univ. of Pennsylvania, Philadelphia, PA

Disclosure Block: **J. Park:** None.

â□□Genome-firstâ□□ approaches to analyzing rare variants can reveal new insights into human biology and disease. Because pathogenic variants are often rare, new discovery requires aggregating rare coding variants into â□□gene burdensâ□□ for sufficient power. However, a major challenge is deciding which variants to include in gene burden tests. Pathogenic variants in *MYBPC3* and *MYH7* are well-known causes of hypertrophic cardiomyopathy (HCM), and focusing on these â□□positive controlâ□□ genes in a genome-first approach could help inform variant selection methods and gene burdening strategies for other genes and diseases. Integrating exome sequences with electronic health records among 41,759 participants in the Penn Medicine BioBank, we evaluated the performance of aggregating predicted loss-of-function (pLOF) and/or predicted deleterious missense (pDM) variants in *MYBPC3* and *MYH7* for gene burden phenome-wide association studies (PheWAS). The approach to grouping rare variants for these two genes produced very different results: pLOFs but not pDM variants in *MYBPC3* were strongly associated with HCM, whereas the opposite was true for *MYH7*. Detailed review of clinical charts revealed that only 38.5% of patients with HCM diagnoses carrying an HCM-associated variant in *MYBPC3* or *MYH7* had a clinical genetic test result. Additionally, 26.7% of *MYBPC3* pLOF carriers without HCM diagnoses had clear evidence of left atrial enlargement and/or septal/LV hypertrophy on echocardiography. Our study shows the importance of evaluating both pLOF and pDM variants for gene burden testing in future studies to uncover novel gene-disease relationships and identify new pathogenic loss-of-function variants across the human genome through genome-first analyses of healthcare-based populations.

PrgmNr 1170 - Incomplete penetrance of disease variants in the UK Biobank

View session detail

Author Block: **A. Barton**[1,2,3], M. Hujoel[1,2], M. A. Sherman[4,1,2], R. E. Mukamel[1,2], P-R. Loh[1,2]; [1]Div. of Genetics, Dept. of Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Bioinformatics and Integrative Genomics Program, Dept. of BioMed. Informatics, Harvard Med. Sch., Boston, MA, [4]Computer Sci. and Artificial Intelligence Lab., Massachusetts Inst. of Technology, Cambridge, MA

Disclosure Block: **A. Barton:** None.

Recent work has found increasing evidence of mitigated, incompletely penetrant phenotypes in heterozygous carriers of disease-causing recessive variants. We leveraged whole-exome imputation within the UK Biobank cohort (*N*~500K) to extend such analyses to 3,481 variants (mostly very rare; median MAF=3.3 x 10^{-5}) annotated as â□□pathogenicâ□□ or â□□likely-pathogenicâ□□ in ClinVar and as â□□autosomal recessiveâ□□ in OMIM.

Testing these variants for association with 54 quantitative traits yielded 98 significant (*p* -8) associations spanning 38 traits and involving variants previously implicated in 33 different diseases. Notable examples included a *POR* missense variant implicated in Antley-Bixler syndrome that associated with a 0.27 (s.e. 0.04) SD increase in height, and an *ABCA3* missense variant implicated in interstitial lung disease that associated with a -0.12 (0.01) SD change in FEV1/FVC ratio (a measure of lung function). Association analyses with 1,257 disease traits yielded 16 additional variant-disease associations (*p* -8). One such association, involving a splice donor variant in *IFT140* (which encodes a protein involved in ciliary function), appeared to be novel. This variant, which has been implicated in retinitis pigmentosa and short-rib thoracic dysplasia, associated with cystic kidney disease in heterozygous carriers (OR = 18.4 (8.5-40.0)).

To explore potential mechanisms that might underlie incomplete penetrance, we examined the â□□modified penetranceâ□□ model of Castel *et al*. (2018) proposing that phenotypic effects of loss-of-function (LoF) variants may be modulated by *cis*-eQTLs regulating the expression of the functional allele on the opposite haplotype. We were powered to explore this model at *FLG*, for which 10% of UK Biobank participants carry a LoF variant associated with asthma or atopic dermatitis. However, association analyses in heterozygous carriers (restricting to variants on the functional haplotypes) did not detect any Bonferroni-significant associations with risk of either disease. As an alternative way to investigate this hypothesis, we also analyzed quantitative traits for which we could identify both large-effect coding variants and cis-eQTLs strongly associated with the same trait (*p* -50), searching for differential phenotypic effects depending on the relative phase of such variant pairs. These analyses also did not find any significant interaction effects; however, power was limited due to small numbers of carriers.

Our results show that many disease-associated recessive variants can produce mitigated phenotypes in heterozygous carriers and motivate further work exploring penetrance mechanisms.

PrgmNr 1171 - Biased pathogenic assertions of loss of function variants challenge molecular diagnosis of admixed individuals

View session detail

Author Block: **M. S. Naslavsky**[1,2], M. O. Scliar[1], K. Nunes[3], J. T. Y. Wang[1], G. L. Yamamoto[4], H. Guio[5], E. M. Tarazona-Santos[6], Y. A. O. Duarte[1], M. R. Passos-Bueno[7], D. Meyer[1], M. Zatz[8]; [1]Univ. of Sao Paulo, Sao Paulo, Brazil, [2]Hosp. Israelita Albert Einstein, Sao Paulo, Brazil, [3]Univ.e de SÃ£o Paulo, Sao Paulo, Brazil, [4]Inst. da CrianÃ§a - Univ. of SÃ£o Paulo, SÃ£o Paulo, Brazil, [5]INBIOMEDIC, Lima, Lima, Peru, [6]Univ Federal Minas Gerais, Belo Horizonte, Brazil, [7]Univ De Sao Paulo, Sao Paulo, Brazil, [8]Univ.e de Sao Paulo, Sao Paulo, Brazil

Disclosure Block: **M.S. Naslavsky:** None.

Genomic analyses for diagnostic purposes of individuals affected by monogenic disorders was significantly improved by next-generation sequencing targeting clinically relevant genes. Whole exomes yield a large number of variants that require several filtering steps, prioritization, and pathogenicity classification. Among the criteria recommended by the American College of Medical Genetics and Genomics (ACMG), those that rely on population databases critically affect analyses of individuals with underrepresented ancestries. Population-specific allelic frequencies need consideration when characterizing potential deleteriousness of variants. An orthogonal input for classification is annotation of variants previously classified as pathogenic as a criterion that provide supporting evidence widely sourced at ClinVar. We used a whole-genome dataset from a census-based cohort of 1,171 elderly individuals from SÃ£o Paulo, Brazil, highly admixed and unaffected by severe monogenic disorders, to investigate if pathogenic assertions in ClinVar are enriched with higher proportions of European ancestry, indicating bias. Potential loss of function (pLOF) variants were filtered from 4,250 genes associated with Mendelian disorders and annotated with ClinVar assertions. Over 1,800 single nucleotide pLOF variants were included, 381 had non-Benign assertions. Among carriers (N=463), average European ancestry was significantly higher than non-carriers (N=708; p=0.011). pLOFs in genomic contexts of non-European local ancestries were nearly three times less likely to have any ClinVar entry (OR=0.353;p

PrgmNr 1172 - Leveraging human phenotype ontology based similarity to further functional annotation of the human genome

View session detail

Author Block: **A. Jolly**[1,2], H. Du[1], M. Dawood[3,2], Z. Coban-Akdemir[4,1], P. Liu[1], R. A. Gibbs[5], J. E. Posey[1], J. R. Lupski[1,6,7,3]; [1]Baylor Coll. of Med., Houston, TX, [2]BCM Med. Scientist Training Program, Houston, TX, [3]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [4]The Univ. of Texas Hlth.Sci. Ctr., Houston, TX, [5]Baylor Coll. Med., Houston, TX, [6]Texas Children's Hosp., Houston, TX, [7]Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX

Disclosure Block: **A. Jolly:** None.

Next generation sequencing has contributed to the functional annotation of ~6000 computationally annotated genes mapped on the haploid human genome reference. Exome sequencing (ES) studies within the Centers for Mendelian Genomics (CMG) have driven discovery at a continuous rate of 1 per 30 ES and revealed aspects of molecular pathobiology in human disease. The discovery of previously unpublished gene-disease associations has been shown to be the most contributory factor to an increased molecular diagnostic rate upon ES reanalysis (Liu, *et al*. 2019. *NEJM*. *380(*25*)*:2478-2480). Moreover, use of the American College of Medical Genetics and Genomics criteria to classify pathogenic variants has allowed for the definitive demonstration of multi-locus pathogenic variation (MPV), i.e., multiple molecular diagnoses, in ~5% of clinical ES cases. Despite these advances, there remains a large proportion of cases unsolved by ES or whole genome sequencing (WGS), a substantial fraction of which we hypothesize are due to as-yet unrecognized Mendelian disease genes. Previously published exome reanalysis pipelines leveraged the Human Phenotype Ontology (HPO) to match known disease associated genes with a proband phenotype. We sought to expand these approaches by using HPOFiller to annotate non-disease associated genes with HPO terms to inform novel gene-phenotype associations. HPO annotated disease associated genes and non-disease associated genes were used to query a phenotypic match to probands/cohorts within the Baylor Hopkins CMG (BHCMG) database. Proof of concept studies successfully allowed for i) determination of individual variant impact within a case of MPV; ii) dissection of allele specific phenotypic spectrum within a genetically heterogeneous skeletal dysplasia cohort (Robinow Syndrome); iii) identification of critical dosage sensitive genes driving phenotype manifestation in a case of multiple *de novo* CNV; and iv) elucidation of the phenotypic spectrum of a gene associated with a birth defect sequence affecting the genitourinary tract. We now apply this discovery approach to the OMIM disease HPO database in order to cluster disease diagnoses by their phenotypic features, i.e. clinical synopsis of associated trait, rather than subjective clinical categorization in order to achieve a molecularly based categorization of congenital disease. Furthermore, we apply these novel experimental genomics and computational analytic approaches to the whole BHCMG database (~12,000 exomes) to solve previously unsolved cases and elucidate the prevalence of MPV across multiple disease phenotypes in the research sequencing setting.

PrgmNr 1175 - Attitudes in four Latin American countries towards donation of personal genetic data for research

View session detail

Author Block: **G. Chavarria-Soley**[1], F. Francis-Cartin[1], G. Macaya[2], R. Alfaro-Redondo[1], L. Robarts[3], A. Middleton[3], H. Raventos[4]; [1]Univ. de Costa Rica, San JosÃ©, Costa Rica, [2]Univ. de Costa Rica, San Pedro de Montes de Oca, San Jose, Costa Rica, [3]Wellcome Genome Campus, Cambridge, United Kingdom, [4]Univ Costa Rica, San Pedro, Costa Rica
Disclosure Block: **G. Chavarria-Soley:** None.
Currently, large amounts of genetic and genomic information are being generated worldwide. Making this data available to the research community is an attractive concept with many potential benefits, but which also raises various concerns. Ultimately, it is the person whose genetic data is being generated who decides whether to share it. â□□Your DNA, Your Sayâ□□ is an anonymous online survey that covers attitudes of the public towards genomic and medical data donation, trust in research professionals, and concerns about consequences of reidentification. It has been translated into 15 languages and has been answered by individuals from more than 20 countries. We explore the results of the survey in four Latin American countries: Argentina (N=919), Costa Rica (N=224, a pilot sample), Mexico (N=1347), and Brazil (N=1349). When confronted with different potential uses of their data, the four countries show a similar response pattern. A high proportion of individuals (59.6%-81,7%) are willing to donate their genetic data to their medical doctors, and the numbers are similar (55.9%-86.2%) when the intended users of the data are non-profit researchers. Willingness to donate in these two scenarios is much higher in Costa Rica. This is probably because the pilot sample is small and skewed towards people with higher education. In all countries, however, participants are significantly less likely to donate data to for-profit researchers (33.9%-38.4% willingness to donate). The three most cited concerns regarding donation of their genetic data are the same in Mexico, Argentina and Brazil: â□□my DNA being copied and then planted at the scene of a crimeâ□□, â□□being clonedâ□□, and â□□upsetting my genetic relativesâ□□. Again, the ranking of concerns in Costa Rica differs from the other countries. Participants were asked whether they would trust different persons with their genetic information. The pattern is similar in the four countries, with the participantâ□□s own doctor as the most trusted person (over 68% for all countries), followed by a researcher at a university in the participantâ□□s own country (over 30% for all countries). The lowest trust values were found for researchers at a company worldwide and governments worldwide. The â□□Your DNA, Your Sayâ□□ survey provides a valuable opportunity to study and compare the views on genetic and genomic data sharing in different regions of the world. Our results suggest that Latin American countries share similar views on the subject.

PrgmNr 1176 - Patient completion of updated hereditary cancer genetic testing

View session detail

Author Block: K. Clift, **S. K. Macklin-Mantia**, S. L. Hines; Mayo Clinic, Jacksonville, FL
Disclosure Block: **S.K. Macklin-Mantia:** None.
For many years, hereditary cancer genetic testing was often limited to one or two genes (i.e. *BRCA1* and *BRCA2*). Identification of additional hereditary cancer genes and advancements in testing technology have led to panel testing becoming commonplace. Individuals with previous nondiagnostic results may have a pathogenic variant in a more newly characterized hereditary cancer gene, and they and their family members may benefit greatly from recontact and updated expanded panel testing. Our aim was to identify how well follow up for updated analysis had been completed and review results of such expanded testing. From 2003-2020, 708 patients were identified who had completed *BRCA1/2* only analysis. The majority had an initial nondiagnostic result (674). Only about 15% (109) had already completed additional, panel genetic testing; 15 (13.7%) were found to carry a pathogenic variant. This study identified 368 patients with a nondiagnostic result who had not completed additional testing, were not known to be deceased, lived within the U.S., and had an email address within the medical record system. A recruitment message was sent to patientsâ email addresses and through their patient portals. An at-home salvia kit for a multi-cancer panel of 84 genes was ordered for participants at no personal charge, and no in person visit was required for testing. Of the 80 preliminary results thus far, 6 patients carried a pathogenic variant (7.5%), specifically in *ATM*, *CHEK2*, *MUTYH* (x2), *PALB2*, and *RAD50*. Thirty-nine patients carried one or more variant(s) of uncertain significance; 35 patients received negative results. Almost 1/3 (N=100, 27%) were noted to have opened their portal communication and then not pursue the option for complimentary, updated testing. These results further underscore the need to recontact patients to identify families with previously undetectable genetic risk for hereditary cancers to allow for individualized management. Many patients who would benefit from updated hereditary cancer genetic testing have not completed and chose not to pursue testing even though many potential barriers, such as need for travel, access to a genetics provider, and cost, were removed. Further understanding into what drives these decisions is warranted.

PrgmNr 1177 - Awareness and Use of Genetic Testing: An Analysis of the Health Information Trends Survey 2020

View session detail

Author Block: **J. Tiner**, L. E. Mechanic, L. Gallicchio, E. M. Gillanders, K. Helzlsouer; DCCPS/EGRP, Natl. Cancer Inst., Rockville, MD
Disclosure Block: **J. Tiner:** None.
Background: The growing availability and commercialization of genetic tests has improved public knowledge and uptake. Genetic testing for clinical and recreational purposes (e.g. ancestry testing) has increased in recent years. However, not all groups have had equal access and opportunity to use the technology. Previous research suggests that there are disparities in awareness and implementation of genetic testing specifically for racial and ethnic minorities and those with low income or education. **Goals**: This project examined awareness and use of genetic testing and associations with demographic characteristics and cancer history. **Methods**: The Health Information National Trends Survey (HINTS) is a national survey that regularly collects information on public awareness and use of cancer- and health-related information. Using HINTS 5 Cycle 4 data, collected in 2020, we examined responses to questions asking participants if they had ever heard of or ever had genetic testing. We examined the associations between demographic characteristics and ever heard of or had genetic testing using chi square tests. **Results:** A total of 3,865 participants completed the HINTS questionnaire for a response rate of 38%. Of these, 3,767 responded to the genetics related questions and were included in this analysis. Overall, 75% of participants reported that they had heard of genetic testing and 19% of participants reported they had genetic testing. Ancestry testing was the most common type of genetic test either heard of or received. Fourteen percent of those with a history of cancer reported having cancer related genetic testing. Of those with a history of cancer, 18% reported a history of breast cancer and 5% reported a history of colon cancer. Ethnicity, education, income, marital status and having a family history of cancer were associated with both hearing of and having any type of genetic test (pConclusion: The results of this analysis align with previously published work showing racial, ethnic, income and education disparities in knowledge and use of genetic testing. While few respondents who had a history of cancer reported having had cancer-related genetic testing, more details are needed to determine the appropriateness or eligibility for testing. These differences in knowledge and use of genetic testing support the need for improved communication and increased engagement within these communities.

PrgmNr 1178 - Health Locus of Control, Belief in Genetic Transmission of COVID-19, and COVID-related Outcomes in a Global Sample

View session detail

Author Block: **T. D. Dye**, S. Siddiqi, E. Pressman; Univ. of Rochester Sch. of Med. and Dentistry, Rochester, NY
Disclosure Block: **T.D. Dye:** None.
Perception that a disease is transmitted through â□□geneticsâ□□ can result in inaction and unnecessary exposure if genetically-transmitted diseases are perceived (as they often are) as untreatable/not preventable. We sought to examine how health locus of control was associated with belief in genetic transmission of COVID-19 (a viral disease not transmitted genetically), and how belief in COVID-19 genetic transmission related to a range of social and health outcomes. **Methods**: We ascertained belief in COVID-19 transmission modalities in a global study of COVID-19-related experience in April/May 2020 (early in the pandemic). In total, 7,411 respondents from 175 countries were recruited through social media and Amazonâ□□s mTURK digital workforce. Multidimensional Health Locus of Control (MHLC â□□Powerful Others,â□□ â□□Chance,â□□ â□□Internalâ□□) reflected oneâ□□s belief in influences on their health. We used logistic regression to generate adjusted Odds Ratios (aOR) and 95% Confidence Intervals (95% CI) to ascertain magnitude and significance of the association (<.05 between mhlc and belief in genetic transmission subsequently on covid-19 social health outcomes.>Results: In total, 7.1% (n=525) of participants believed that COVID-19 was transmitted through genes, with another 12.2% (905) unsure. Adjusting for age and education, participants with higher levels on all three MHLC domains were significantly more likely to believe COVID-19 was transmitted through genes (or were unsure): (MHLC-Chance aOR: 2.47; 95% CI: 2.13, 2.87; MHLC-Internal aOR: 1.50; 95% CI: 1.30, 1.73; MHLC-Powerful Others aOR: 2.51; 95% CI: 2.17, 2.90). Controlling for age, education, and all three MHLC levels, belief that COVID-19 was transmitted by genes was significantly associated with self-reported COVID-19 infection (aOR: 1.57; 1.24, 1.98), greater COVID-19-related worry (aOR: 1.16; 1.00, 1.35), vaccine hesitancy (aOR: 1.57; 1.34, 1.83), and higher COVID-19 livelihood impact (aOR: 1.53; 1.32, 1.78). **Discussion**: Early in the COVID-19 pandemic, belief that the virus was genetically transmitted may have influenced behavior to place individuals at excess risk for infection and poor COVID-19 outcomes, especially among people with stronger beliefs in the effects of fate and â□□powerful othersâ□□ on their health. Misunderstanding genetic transmission of disease - as in the case of COVID-19 - could have adverse health consequences in clinical practice and public health. Better understanding perceptions of genetic transmission and of COVID-19 may help reduce misunderstanding, improve perceptions and knowledge of treating COVID-19, and promote preventive behaviors in a more timely manner.

PrgmNr 1179 - Identifying message design considerations to motivate cascade testing for familial hypercholesterolemia: Stakeholders' advice for healthcare providers

View session detail

Author Block: **G. Campbell-Salome**[1], W. L. Nicole[2], I. G. Ladd[2], A. Sheldon[3], C. Davis Ahmed[3], A. Brangan[2], M. N. McMinn[2], A. K. Rahm[2], M. L. Schwartz[2], E. Tricou[2], C. L. Fisher[1], A. C. Sturm[4]; [1]Univ. of Florida, Gainesville, FL, [2]Geisinger, Danville, PA, [3]The FH Fndn., Pasadena, CA, [4]Geisinger, Lewis Center, OH

Disclosure Block: **G. Campbell-Salome:** None.

Purpose: Motivating families with familial hypercholesterolemia (FH) to undergo cascade testing is critical for diagnosis and life-saving treatment. Healthcare providers (HCPs) can assist in family communication about FH and motivate cascade testing uptake as authoritative sources of information. Identifying how to strategically communicate about FH risks can guide HCPs on effectively motivating at-risk relative to pursue cascade testing. The Extended Parallel Process Model (EPPM) was used to identify message approaches that focus on four key factors (severity, susceptibility, response efficacy, self-efficacy) that predict motivation to engage in a health behavior. The EPPM is a well-supported, contemporary model for strategic risk communication and has been applied to create messages promoting health protective behaviors in areas of smoking cessation, heart disease, and cancer screening. This study gathered stakeholdersâ perspectives on how to craft messages from HCPs to promote cascade testing uptake, informed by the EPPM. Methods: Participants with FH were recruited from a healthcare system in Pennsylvania, USA, and The FH Foundation. Participants were asked to recruit a family member for a dyadic interview. Additional participants responded to a survey. Open-ended survey and interview data captured stakeholdersâ feedback on written and verbal messages that were used to facilitate family communication about FH and cascade testing. Data were thematically analyzed and triangulated. Results: Participants included 22 individuals in 11 dyadic interviews and 98 survey respondents. Message approaches that address EPPM factors were identified to motivate cascade testing. Recommendations highlight the importance of balancing relativesâ sense of threat (severity and susceptibility) with their ability to reduce risk (efficacy). For *severity*, HCPs should list health risks associated with untreated FH and emphasize FH as a genetic condition affecting generations. For *susceptibility*, HCPs should explain how FH is inherited and that young relatives are susceptible. To promote *response efficacy*, HCPs should describe testing as providing a definitive FH diagnosis, highlight FH as easily treatable, and stress the value of early FH intervention. To promote *self-efficacy*, HCPs should clarify testing options and costs, provide instructions on how to pursue testing, and offer support resources. Conclusions: Results highlight ways to design messages to motivate cascade testing for FH and other hereditary conditions. A prospective, pragmatic trial will test these approaches to examine the impact of HCPs messages on cascade testing uptake.

PrgmNr 1180 - Usual care genetic counseling vs. ARIA: An RCT of an Accessible, Relational, Inclusive and Actionable model for sequencing results disclosure to historically underrepresented populations in the CSER CHARM study

View session detail

Author Block: **G. Joseph**[1], M. C. Leo[2], B. B. Biesecker[3], L. Riddle[1], C. Guerra[1], L. Amendola[4], M. J. Gilmore[2], J. M. Zepp[2], B. Rolf[5], K. P. Anderson[6], S. Okuyama-Sasaki[6], M. Caruncho[1], J. E. Hunter[2], E. M. Keast[2], T. L. Kauffman[2], N. M. Lindberg[2], A. A. Reyes[2], G. P. Jarvik[7], B. S. Wilfond[8], K. A. Goddard[9]; [1]Univ California San Francisco, San Francisco, CA, [2]Kaiser Permanente Ctr. for Hlth.Res., Portland, OR, [3]RTI Intl., Washington, DC, [4]Univ. of Washington, Seattle, WA, [5]Univ Washington, Seattle, Seattle, WA, [6]Denver Hlth.Ambulatory Care Services, Denver, CO, [7]Univ Washington Med Ctr., Seattle, WA, [8]Seattle Children's Res. Inst., Seattle, WA, [9]Kaiser Permanente Northwest, Portland, OR
Disclosure Block: **G. Joseph:** None.
Background. Barriers to effective communication in genetic counseling include high oral literacy demand, dominance of informational vs. psychosocial counseling dialogue, and limited provision of valued information. Prior research with historically underserved patients found that ineffective communication contributed to poor understanding and limited engagement. Methods. Within the NIH CSER consortium CHARM study, which recruited historically underserved populations, we assessed ARIA (Accessible, Relational, Inclusive and Actionable), an evidence-based genetic counseling communication model. ARIA is designed to increase accessibility of genetic counseling (GC) for all patients regardless of prior genetics knowledge or literacy level, and requires in-depth training of GCs in specific communication skills. We conducted a pragmatic randomized controlled trial to compare usual care (UC) GC to ARIA for the return of exome sequencing results in participants ages 18-49. We hypothesized that ARIA would be more effective in achieving satisfaction with communication and increasing essential knowledge. Two GCs were assigned to each arm for the duration of the trial. Study participants were randomized 1:1 and were blinded to the type of counseling they received. Results. Participants (353 ARIA; 344 UC) had a mean age of 34.6 (SD=8.2); 71% were female, 44% were low income, and 15% preferred Spanish. Almost 20% had marginal or inadequate health literacy. Preliminary analyses indicate that participants in both arms were highly satisfied, with high agreement on being treated with sensitivity and respect (100% UC vs 99% ARIA), feeling listened to (99% UC vs 98% ARIA), getting clear, understandable information (98% UC vs ARIA 97%), and feeling comfortable asking questions and voicing concerns (96% in UC vs 98% in ARIA). 97% of participants with normal cancer results recalled their results correctly vs. 90% with VUS results vs. 89% with abnormal results. 90% with normal additional finding results were accurate while patients with abnormal additional findings were more likely to recall results correctly if the result was medically actionable (88%) than if not (17%). Mean cancer genetics knowledge for both groups was high at baseline (81% accuracy UC, 78% ARIA) and showed a small increase from baseline to follow-up for both groups (+3% UC, +4% ARIA). Conclusions. Communication satisfaction appears to be similar between arms; inferential testing of the outcomes is in progress. Although knowledge was high, participantsâ recall of their results appeared to be similar between arms but may have varied by type and actionability of the result.

PrgmNr 1183 - Identification of clinically relevant variants in homologous regions in 41,755 exomes

View session detail

Author Block: **W. Steyaert**[1], M. Steehouwer[1], L. Haer-Wigman[1], R. Pfundt[1], A. P. Stegmann[2], H. Yntema[1], H. Brunner[1], A. Hoischen[1], C. Gilissen[1]; [1]Radboud Univ. Nijmegen Med. Ctr., Nijmegen, Netherlands, [2]Univ Hosp Maastricht, Maastricht, Netherlands
Disclosure Block: **W. Steyaert:** None.
Despite routine clinical exome sequencing, the molecular cause of disease remains unknown in the majority of patients with a rare genetic disease. A possible contributor to this problem is the difficulty in analyzing genomic regions with a high sequence similarity to other regions in the genome, i.e. homologous regions. Altogether, these segments comprise about 3.5% of the exome and consist mainly of protein coding genes having one or multiple pseudogenes. Short-read sequencing reads originating from these regions will either align with very low quality or discordant to their biological origin causing most single nucleotide variants (SNVs) as well as copy number variants (CNVs) to be undetected.
To address this problem we devised an algorithm and implemented it in a software package called Cameleolyser. In addition to aligning sequencing reads to a classical reference genome, Cameleolyser also maps reads to a masked version of it. Gene conversions and deletions are identified by using sequence differences between the homologs. To determine SNVs not introduced by a conversion event, a sensitive variant calling is conducted onto the masked alignment.
The application of Cameleolyser to a cohort of 41,755 whole exome sequencing samples yielded 17,394 rare homozygous deletions and 314,131 rare SNVs in these challenging regions. By generating PacBio high-fidelity high coverage genomes for a subset of 20 samples we could confirm 77 out of the 83 SNVs (93%). The SNVs that are the consequence of a gene conversion event (n=8) could all be confirmed. Out of the 15 CNVs that were found by our method in this subset of samples, 11 (73%) corresponded to the PacBio data. Furthermore, 7 randomly chosen ultra-rare SNVs were selected for validation and confirmed by Sanger sequencing.
We then focused on variation that could potentially be disease-causing. In the cohort we found 1,145 ultra-rare homozygous deletions in OMIM genes, 1,011 heterozygous loss-of-function (LoF) variants, and 54 homozygous LoF variants of which 50 due to gene conversions. The top ranked gene is *STRC*, encoding for stereocilin and a known autosomal recessive disease gene for deafness. In *STRC*, we identified 25 homozygous LoF mutations due to gene conversions as well as 37 homozygous deletions. All of these events were exclusively found in patients with hearing impairment.
In conclusion, we developed a novel method that can accurately identify clinically relevant copy number and single nucleotide variations in homologous coding regions from diagnostic exome sequencing data. We demonstrate that disease causing LoF and other variation is often introduced in these regions by means of gene conversions.

PrgmNr 1184 - Analytical validation of clinical whole genome sequencing across a diverse mutation spectrum

View session detail

Author Block: **K. V. Schulze**[1], J. Sinson[1], B. Yuan[2,3,1], V. T. Nguyen[3], M. Santana[3], D. C. Chen[4], C. Qu[3], H. DAI[4,3], C. Shaw[1,3], C. M. Eng[1,3], P. Liu[1,3]; [1]Baylor Coll. of Med., Houston, TX, [2]Seattle Children's Hosp., Seattle, WA, [3]Baylor Genetics Lab., Houston, TX, [4]Baylor Coll. Med., Houston, TX
Disclosure Block: **K.V. Schulze:** None.
Clinical whole genome sequencing (WGS) offers the opportunity to not only end but also prevent many diagnostic odysseys. Our experience suggests that WGS is preferentially ordered for patients with a history of negative molecular or cytogenetic testing, and less frequently requested as a comprehensive first-tier test for patients with a broad differential diagnosis.
To evaluate how well WGS performs compared to other molecular or cytogenetic tests that are commonly used as first-tier diagnostic assays we applied WGS to >250 de-identified cases with a variety of known disease-causing genetic variants that were previously detected in clinical testing using traditional diagnostic methods. For this analytical validation all samples were processed and analyzed in a CLIA/CAP certified laboratory, sequenced with a minimum average coverage of 40X, and their data mapped to the GRCh38 human reference genome using Illuminaâ□□s DRAGEN pipeline. Our study cohort was carefully selected to largely consist of cases with genetic variants that could pose diagnostic challenges for WGS, such as copy number variants (CNVs), in particular those mediated by repetitive elements, structural variants (SVs), such as translocations and mobile element insertions, as well as trinucleotide repeat (TNR) expansions.
In addition to single nucleotide variants that were confirmed both in the nuclear as well as the mitochondrial genome, regions with absence of heterozygosity, and uniparental disomy, we detected 264/265 (99.6%) of reported CNVs from 174 individuals that ranged in size from 0.002-35.1 Mb using a read depth-based algorithm. Our WGS analysis of TNR expansions in 45 cases, of which 26 had repeat lengths in the normal range, showed a sensitivity of 73.7% and specificity of 100% when we labeled samples by mutation categories according to disease-specific thresholds of repeat lengths. However, we noticed that the degree of deviation between repeat lengths found with WGS and those clinically reported drastically grew with increasing TNR sizes; repeat lengths found by WGS did not exceed 150, even for a TNR expansion in the *FXN* gene with a reported approximate size of 1,400. At this stage of our analysis we conclude that WGS can detect many, but not all, TNR expansions. Nonetheless, compared to traditional tests, WGS can reliably identify a diverse spectrum of genetic and genomic variants making it a robust assay suitable for both first- and last-tier diagnostic testing.

PrgmNr 1185 - Genome sequencing as a first-line genetic testing in ill newborns

View session detail

Author Block: **M. L. Thompson**[1], K. M. Bowling[1], D. R. Latner[1], C. R. Finnila[1], S. Hiatt[2], M. Amaral[3], K. M. East[1], W. V. Kelley[1], M. E. Cochran[1], V. Greve[1], J. M. J. Lawlor[1], H. Meddaugh[4], A. Luedecke[5], K. Jackson[6], L. G. Hendon[7], H. M. Janani[8], M. Johnston[4], L. Merin[5], S. L. Deans[6], H. Williams[7], K. Laborde[8], M. B. Neu[9], J. Patrick-Esteve[4], K. B. Brothers[6], R. D. Savich[7], S. Spedale[8], G. S. Barsh[1], B. R. Korf[10], G. M. Cooper[1]; [1]Hudson Alpha Inst. for Biotechnology, Huntsville, AL, [2]601 Genome Way, Huntsville, AL, [3]HudsonAlpha Inst. for Biotechnology, Huntsville, AL, [4]Children's Hosp. New Orleans, New Orleans, LA, [5]Univ. of Alabama at Birmingham, Birmingham, AL, [6]Univ. of Louisville, Louisville, KY, [7]Univ. of Mississippi Med. Ctr., Jackson, MS, [8]Woman's Hosp., Baton Rouge, LA, [9]Huntsville, FL, [10]Univ Alabama at Birmingham, Birmingham, AL

Disclosure Block: **M.L. Thompson:** None.

In SouthSeq, we conducted genome sequencing (GS) as a first-line test for infants with symptoms suggestive of a genetic disorder being cared for in a neonatal intensive care unit (NICU) and other pediatric units. Study recruitment targeted diverse populations representing racial/ethnic minorities and rural medically underserved areas that are historically under-represented in genomic medicine research. Genome sequencing and analysis were performed concurrent with standard clinical care for 367 affected newborns to detect disease-causal genetic variation. Definitive diagnostic (DD) and likely diagnostic (LD) findings were identified in 109 newborns (30%) with an additional 51 (14%) found to harbor an uncertain result. Thirty-nine percent of GS-detected DD/LD findings were identified via standard, concurrent clinical genetic testing, whereas 53.2% were not (7.3% of GS-detected DD/LD findings did not receive any clinical genetic testing), suggesting that GS testing is better for obtaining early genetic diagnosis in newborns. In most cases, GS detected a SNV/indel when only clinical CNV testing was ordered. We also identify phenotypes that correlate with the chances of receiving a DD/LD finding, such as an enrichment for craniofacial, ophthalmologic and auditory abnormalities. We did not observe any differences in diagnostic rates between racial/ethnic minority groups. SouthSeq has provided an early genetic diagnosis and has demonstrated the utility of using GS as a comprehensive first-line genetic test.

PrgmNr 1186 - Genome-to-Treatment: A virtual, automated system for population-scale diagnosis and acute management guidance for genetic diseases in 13.5 hours

View session detail

Author Block: **S. F. Kingsmore**[1], M. J. Owen[1], S. Lefebvre[2], C. Hansen[1], C. Kunard[3], B. R. Lajoie[4], T. McPhail[3], S. Mehtalia[5], D. Dimmock[6], L. D. Smith[7], G. H. Scharer[8], R. R. Mardach[9], M. J. Willis[10], A. S. Feigenbaum[5], S. Rego[1], J. Perry[11], K. Chau[1], L. Van Der Kraan[1], Y. Ding[1], Y. Kwon[1], S. Batalov[12], A-K. Niemi[13], M. Speziale[11], M. S. Wright[14], T. Defay[2], J. Reynders[2], J. Lenberg[1], C. A. Hobbs[15], K. Hall[16], RCIGM Investigators; [1]Rady Children's Inst. for Genomic Med., San Diego, CA, [2]Alexion Pharmaceuticals, Boston, MA, [3]Illumina Inc, San Diego, CA, [4]Illumina Inc., San Diego, CA, [5]San Diego, CA, [6]Rady Children's Inst., Genomic Med., San Diego, CA, [7]Children's Mercy Hosp, Kansas City, MO, [8]RPRD Diagnostics Lab., Milwaukee, WI, [9]Rady Children's Hosp. of San Diego, San Diego, CA, [10]La Jolla, CA, [11]Rady Children's Hosp., San Diego, CA, [12]Rady Childrens Inst., Genomic Med., San Diego, CA, [13]Stanford Univ, Stanford, CA, [14]Rady Children's Inst. for Genomics Med., San Diego, CA, [15]Univ. of Arkansas for Med. Sci., Little Rock, AR, [16]Illumina Cambridge Ltd., Saffron Walden, Essex, United Kingdom

Disclosure Block: **S.F. Kingsmore:** None.

Many rare genetic diseases have effective treatments. Without immediate implementation many progress rapidly to severe morbidity or mortality. Front-line physicians are often unfamiliar with these diseases or treatments. Hence, timely molecular diagnosis may not always improve outcomes. We describe Genome-to-Treatment (GTRx), an automated, virtual system for genetic disease diagnosis and acute management guidance for ill children in intensive care units (ICU). Diagnosis was achieved in 13.5 hours by sequencing library preparation directly from blood, faster whole genome sequencing (WGS) and informatic analysis, natural language processing of electronic health records and automated interpretation. 563 severe genetic diseases with effective treatments were identified by literature review, clinician nomination and WGS experience. Specific treatments - drugs, devices, diets, and surgeries - were identified for each by automated internet and literature searches, and manually curated. Five clinical geneticists adjudicated the indications, contraindications, efficacy, and evidence-of-efficacy of each treatment in each disorder in this clinical context. After discussion, they agreed upon 189 of the first 190 treatments. We integrated 10 genetic disease information resources, and electronically linked them and the adjudicated treatments to each automated diagnostic result (http://gtrx.rbsapp.net/). This system had superior analytic performance for single nucleotide, insertion-deletion, structural and copy number variants. It provided correct diagnoses and acute management guidance in four retrospective patients. Prospectively, an infant with encephalopathy was diagnosed in 13.5 hours, received effective treatment immediately, and had a good outcome. GTRx will facilitate broad implementation of optimal acute treatment for children with rapidly progressive genetic diseases by front-line ICU physicians.

PrgmNr 1187 - Large scale Cas9 mediated depletion of highly abundant transcripts to expand the interpretable genome and improve the diagnostic yield of clinical RNA Seq

View session detail

Author Block: **A. Y. Huang**[1], A. Siddique[2], D. Ramachandran[2], H. Lee[3], L-K. Wang[1], K. Brown[2], S. F. Nelson[4,1]; [1]California Ctr. for Rare Diseases, David Geffen Sch. of Med., Univ. of California-Los Angeles, Los Angeles, CA, [2]Jumpcode Genomics, San Diego, CA, [3]3billion, Seoul, Korea, Republic of, [4]Dept. of Human Genetics, David Geffen Sch. of Med., Univ. of California-Los Angeles, Los Angeles, CA
Disclosure Block: **A.Y. Huang:** None.
Whole genome sequencing (WGS) provides about a 100-fold increase in rare variants identified per human compared to whole exome sequencing (WES). However, recent analyses demonstrate this increased coverage of the genome provides only a modest improvement in diagnosis for rare genetic disorders. This is largely because the functional consequence for most of the non-coding variants identified by WGS remains unknown. To address this, several groups have performed whole-transcriptome sequencing (RNA-Seq) with WGS to improve clinical interpretation of these many variants. Using this combined approach, a 15-30% improvement in overall diagnostic yield has been reported, largely through the identification of rare DNA variants which are demonstrated to cause aberrant mRNA splicing. However, the utility of RNA-Seq is limited to genes that are well-expressed in tissues that are possible to access from patients. Although about 80% of all known disease-associated genes (OMIM) are observed in RNA-Seq from fibroblasts or blood, the vast majority of these are expressed at levels too low for comprehensive analysis of rare splicing variation, and only about 20% of all genes are robustly observed. To address this issue, we have developed a novel method to improve coverage of low-expressed genes through large-scale, selective depletion of the most abundant transcripts (TPM > 30). The CRISPRclean method utilizes Cas9 nuclease and 360,000 guide RNAs to specifically remove RNA-Seq library fragments from over 4,000 targeted genes, which are the well observed genes from non-depleted RNA-Seq, and results in an average six-fold increase in coverage of untargeted genes compared to untreated RNA-Seq libraries. We demonstrate expansion of the number of genes that can be assessed for rare splicing aberrations, or allelic alterations in expression, and thus greatly expand the proportion of the genome interpretable by RNA-Seq. The method can be applied to any cell type of RNA, and can be customized for maximal effect. This depletion method occurs post-library generation, and thus may be applied to already sequenced samples or added to existing clinical RNA-Seq workflows and effectively increases the dynamic range of gene expression observation and alternative splicing from various tissues. We demonstrate diagnostic efficacy by applying CRISPRclean to a large cohort of rare Mendelian disease samples from within the Undiagnosed Disease Network.

PrgmNr 1188 - Integration of proteomics with genomics, transcriptomics and phenomics increases the diagnostic rate of Mendelian disorders

View session detail

Author Block: **D. Smirnov**[1,2], R. Kopajtich[2,1], S. L. Stenton[3,2], S. Loipfinger[4], C. Meng[5], I. F. Scheller[6], V. A. Yepez[1,4], M. Gusic[7], R. Berutti[1], C. Ludwig[5], K. Murayama[8], GENOMIT, J. Gagneur[9,1,2], H. Prokisch[10,1]; [1]Inst. of Human Genetics, Technical Univ. of Munich, Munich, Germany, [2]Helmholtz Ctr. Munich, Munich, Germany, [3]Technical Univ. of Munich, Munich, Germany, [4]Dept. of Informatics, Technical Univ. of Munich, Munich, Germany, [5]Bavarian Ctr. for Biomolecular Mass Spectrometry (BayBioMS), Technical Univ. of Munich, Munich, Germany, [6]Inst. of Computational Biology, Helmholtz Zentrum MÃ¼nchen, Munich, Germany, [7]HelmholtzZentrum MÃ¼nchen, Neuherberg, Germany, [8]Ctr. for Med. Genetics, Dept. of Metabolism, Chiba Childrenâ□□s Hosp., Chiba, Japan, [9]Technical Univ. of Munich, Garching, Germany, [10]Helmholtz Zentrum Muenchen, Neuherberg, Germany
Disclosure Block: **D. Smirnov:** None.

Due to the lack of functional evidence, the diagnostic rate of whole exome and genome sequencing (WES/WGS) in Mendelian disorders plateaus at ~50%, increasing by 10-15% when complemented by transcriptomic data (RNA-seq). Here, we demonstrate the diagnostic power of integrating proteomics into a systematic multi-omic approach and provide a blueprint for its implementation in routine clinical practice.

Quantitative tandem mass tag (TMT) labelled proteomics was applied to fibroblasts from 145 individuals, detecting ~8,000 proteins per sample, spanning over half of all Mendelian-disease genes. To detect protein expression levels significantly outside of the physiological range, as evidence of underlying variant pathogenicity, we developed PROTRIDER, a denoising-autoencoder based algorithm controlling for known and unknown sources of proteome-wide variation. This approach was first validated in 22 samples with confirmed pathogenic variants, by protein underexpression in 64%, while the corresponding transcript was reduced only in 18%. We then applied the approach to 21 unsolved cases with prioritised variants of uncertain significance (VUS), validating 14 diagnoses, 67% by aberrant protein- and 24% (5/21) by aberrant RNA-expression.

In search of the diagnosis in the remaining unsolved cases, we called genome-wide significant RNA and protein expression outliers of three classes (RNA-only, protein-only, and RNA-and-protein) across the entire dataset. Outliers were enriched for splice, stop, and frameshift variants, in addition to missense and in-frame indel variants in protein-only outliers. Integrating these outlier calls with rare variant allelic data from WES/WGS and phenotype similarity scores for known disease-gene associations led to the diagnosis of 12 further unsolved cases and identified a novel mitochondrial disease gene, MRPL38.

Overall, our integrative multi-omic approach led to genetic resolution for 21% (26/121) of all unsolved WES/WGS cases and demonstrated the added value of proteomics both in detecting the functional consequence of missense variants, the most frequent VUS, and in delineating the downstream functional consequence of pathogenic variants on the protein complex level (39% of all diagnosed cases). The multi-omics pipeline is available at https://github.com/prokischlab/omicsDiagnostics/ .

PrgmNr 1191 - Integrative analysis of human genetic association studies, single cell transcriptomics, and knockout mouse models identifies cell types and genes involved in skeletal disease

View session detail

Author Block: **M. Lundberg**[1,2], R. C. Chai[3,4], M. Frysz[5], P. Baldock[3,4], W. Khoo[3,4], D. Bassett[6], G. R. Williams[6], P. I. Croucher[3,4], J. P. Kemp[2,5]; [1]The Univ. of Queensland Diamantina Inst., The Univ. of Queensland, Queensland, 4102, Woolloongabba, Australia, [2]Inst. for Molecular BioSci., The Univ. of Queensland, Queensland, 4072, St Lucia, Australia, [3]Garvan Inst. of Med. Res., New South Wales, 2010, Sydney, Australia, [4]St Vincentâ□□s Clinical Sch., Faculty of Med., UNSW Sydney, NSW, 2010, Darlinghurst, Australia, [5]MRC Integrative Epidemiology Unit (IEU) at the Univ. of Bristol, Oakfield House, Oakfield Grove, BS8 2BN, Bristol, United Kingdom, [6]Molecular Endocrinology Lab., Dept. of Metabolism, Digestion and Reproduction, Imperial Coll. London, W12 0NN, London, United Kingdom
Disclosure Block: **M. Lundberg:** None.
Motivation: Genome-wide association studies (GWAS) have identified hundreds of genomic regions associated with skeletal disease traits. However, few of the causal genes underlying these associations have been identified and functionally validated *in vivo.*
Objective: We hypothesised that integration of human genetic association studies and single cell RNA sequencing (scRNA-seq) with phenotype data from knockout (KO) mouse models would identify disease-causing genes and identify the cellular context through which they function.
Methods: scRNA-seq was used to map the transcriptome of cells isolated from the endosteal and marrow compartments of mouse femurs. Hypergeometric tests were used to identify cell types enriched for expression of genes involved in monogenetic skeletal disorders. GWAS, and MAGMA gene-set analysis of 448,010 participants in the UK-Biobank Study was used to identify cell types enriched for bone mineral density (BMD) and height-associated genes. Skeletal phenotyping of >1000 unselected KO mouse lines was used to validate the functional roles of BMD and height-associated genes.
Results: scRNA-seq analysis of 133,942 bone and marrow cells identified 34 distinct cell types including multiple clusters of myeloid, lymphoid, and non-haematopoietic cells, each with distinct transcriptional profiles. Hypergeometric tests showed that osteoblasts and osteoclasts were enriched for human orthologues involved in monogenetic low and high bone mass disorders, respectively (P-7), whereas chondrocytes were enriched for orthologues involved in bone growth (P-9). GWAS of BMD and height, identified 90 novel loci including one intersecting *PLS3*, a known cause of juvenile X-linked osteoporosis (P-9). MAGMA gene set analysis involving GWAS and scRNA-seq results showed that osteoblasts, endothelial and vascular smooth muscle (VSM) cells were enriched for BMD-associated genes (P-4), whereas chondrocytes and B-cells were enriched for height-associated genes (P-3). Eight osteoblast, 7 endothelial and 5 VSM-specific genes associated with BMD were present in a cohort of 1000 unselected KO mouse lines. Deletion of 5, 5 and 4 of the corresponding genes in mice resulted in abnormal skeletal phenotypes. These included novel BMD-associated genes with previously unappreciated roles in the skeleton, including endothelial-specific gene *Slc9a3r2*, which when deleted, resulted in reduced trabecular bone mass.
Conclusions: Our multiscale approach identified novel candidate skeletal disease genes and provides insight into the cellular context through which they may contribute to the development of skeletal disease.

PrgmNr 1192 - A modular massively parallel reporter assay uncovers context-specific allelic activity of GWAS variants

View session detail

Author Block: **A. Tovar**, A. Varshney, Y. Kyono, J. O. Kitzman, S. C. J. Parker; Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **A. Tovar:** None.
Genome-wide association studies have nominated hundreds of genetic signals associated with the prevalent metabolic disease type 2 diabetes (T2D) and related traits. A majority (~90%) of these signals lie in noncoding regions of the genome, suggesting that dysregulated transcriptional circuitry is a primary mechanism contributing to T2D etiology. Predicting functional consequences of noncoding variants remains non-trivial; thus, specific insights about the relationship of these variants to effector transcripts and associated regulatory activity remains limited. To address this outstanding priority, we devised a modified STARR-seq (self-transcribing active regulatory region sequencing) strategy to systematically interrogate enhancer activity of the elements in which these variants reside, and how variants modulate this activity. We constructed a library of 79,380 oligonucleotides spanning the genomic context of 13,230 allelic pairs which we cloned into a vector with a series of configurations to assess position- and promoter-specific regulatory activity. Specifically, we placed inserts both up- and downstream of either a synthetic core promoter (*SCP1*) or a biologically relevant promoter for the human insulin gene (*INS*), then transfected the final library into the 832/13 rat insulinoma cell line to examine effects in pancreatic Î² cells. Overall, around 56% of inserts (n = 44,355/79,380) exhibited significant enhancer activity (FDR INS promoter (n = 370 upstream, 1,251 downstream, 16 common), while only two inserts displayed significant allelic effects across all four configurations. Together, these results indicate that reporter construct design clearly influences assayed reporter activity and hence, use of tissue-relevant promoters involved in disease pathophysiology will be key for comprehensively characterizing the allelic effects of disease-associated variants. Future work will integrate these results with existing genetic and genomic datasets from pancreatic islets and/or Î² cells to explore the regulatory features modified by these variants and prioritize specific elements for follow-up studies.

PrgmNr 1193 - Understanding cell-type-specific drug effects while controlling for genotype

View session detail

Author Block: **F. Wessely**[1], Z. Cader[2], C. Webber[1], StemBANCC consortium; [1]Cardiff Univ., Cardiff, United Kingdom, [2]Univ. of Oxford, Oxford, United Kingdom
Disclosure Block: **F. Wessely:** None.
Cells derived from human induced pluripotent stem cells (iPSCs) offer the possibility to study cell-type-specific drug effects within a controlled in vitro system. An unbiased and informative approach is to measure and compare the cellular transcriptomic response induced upon exposure to the compound. While there are large-scale drug perturbation databases available offering an important resource for drug discovery and repurposing, response profiles across different cell types are usually confounded by differing genotypes. To control for genotype while varying cell type, 12 different cell types were generated from three iPSC-lines derived from healthy and unrelated donors. Differentiated and undifferentiated cells were exposed for seven hours to a selected panel of 12 small molecule compounds with a broad range of mechanisms of action. Transcriptomic responses after compound or control vehicle exposure was measured by 3â mRNA-sequencing yielding approximately 1,700 gene expression profiles. Overall, we observed a highly variable transcriptional response between different cell types treated with the same compound. However, for the majority of compounds a small core set of common response genes shared between eight or more cell types was found, which also contributed to a common early response on the pathway level. This core set can be as specific as a single response gene, highlighting the power for high-resolution platforms to detect very subtle perturbations. Remarkably, even undifferentiated cells often share a common drug response with differentiated cells. And while in most cases variability of the proposed target gene expression was only moderately correlated with the response signal, we identified cases where other modulators such as specific transporters have a large impact on drug response.

PrgmNr 1194 - Modeling eQTLs at single-cell resolution identifies >2000 eQTLs with differential effects across continuous T cell states

View session detail

Author Block: **A. Nathan**[1,2], S. Asgari[1,2], K. Ishigaki[1,2], T. Amariuta[3], Y. Luo[1], J. I. Beynor[1], L. Lecca[4], M. B. Murray[1], D. Moody[1], S. Raychaudhuri[1,2,5]; [1]Harvard Med. Sch./Brigham and Women's Hosp., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [4]Socios En Salud, Lima, Peru, [5]Univ. of Manchester, Manchester, United Kingdom
Disclosure Block: **A. Nathan:** None.
T cells have fluid functional cell states (e.g., cytotoxicity, activation) that we have previously captured at high resolution with canonical correlation analysis (CCA) of single-cell mRNA and protein data (Nathan et al. *Nat Immunol* 2021). These states can trigger autoimmunity or infection response, but to understand genetic effects on such traits, it is essential to define the influence of cell states on gene regulatory variation. However, eQTL studies often obscure high-resolution states by reshaping single-cell data into discrete pseudobulk states. Here, we developed a powerful strategy to define state-dependent eQTLs by modeling the relationship between single-cell gene expression and genetic variants and their interactions with continuous T cell states using a Poisson mixed effects model.We conducted single-cell eQTL analysis of a CITE-seq dataset of over 500,000 memory T cells from an admixed Peruvian cohort to identify 6,511 pan-memory-T cell eQTLs, including previously undescribed eQTLs specific to Peruvian ancestry (e.g., rs9927852 regulating *MAF*). Using CCA-defined T cell states, we modeled eQTL interactions with the linear combination of the top seven continuous axes and found significant state-dependence of over one-third (2,237) of memory-T-cell eQTLs. This enabled us to better identify eQTLs interacting with potentially disease-relevant states like regulatory function: the continuous model not only recapitulated 74% of eGenes with regulatory T cell cluster-dependent effects, but also identified 744 other eGenes associated with continuous degree of regulatory function. Notably, independent eQTLs for the same eGene may have opposing correlations with T cell states, such as for *MDGA1*. After fine-mapping the effect at each locus, state-dependent eQTLs were more enriched in cell-state-specific regulatory regions compared to state-independent eQTLs. We identified many disease-associated variants with state-dependent effects, including autoimmune-associated rs3087243 and *CTLA4*.Our findings show that a single-cell eQTL model is essential to fully leverage single-cell data to study state-specific gene regulation. The regulatory specificity and disease association of these state-dependent eQTLs underscore their biological importance for expanding our understanding of T cell states in health and disease.

PrgmNr 1195 - Sex differences in the expression and genetic regulation of drug metabolism and transporter genes in human liver

View session detail

Author Block: **Y. Huang**[1], F. Li[2], W. Zhang[3], Y. Shan[1], L. Cheung[1], B. E. Stranger[4], R. S. Huang[5]; [1]Univ. of Minnesota, Minneapolis, MN, [2]Mayo Clinic, Rochester, MN, [3]Univ. of Minnesota, Rochester, MN, [4]Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [5]Univ Minnesota, Minneapolis, MN
Disclosure Block: **Y. Huang:** None.
Human metabolic activity and the elimination of xenobiotics exhibit sex differences. However, little is known about the underlying biology of the sex differences. Understanding the underlying cause of these sex differences can provide critical information in improving human health. Here, we characterized the sex differences in the expression level and genetic regulation of drug metabolism and transporter (DMET) genes, which have been well recognized in transforming both endogenous and exogenous compounds and are critical for human health. We first examine the differential gene expression in human liver using transcriptome data in the Genotype-Tissue Expression (GTEx) project. Among 372 DMET genes evaluated, 20 exhibit differences between the two sexes (FDR CYP3A4, *CYP2C19*, *CYP1A2* exhibit sex-biased expression, which cautions the use of medication between males and females. To identify sex-specific genetic effects on the human liver transcriptome, we then conducted the sex-stratified cis-expression quantitative trait loci (cis-eQTLs) analysis. We identified 71 single nucleotide variants (SNPs) that exhibit sex-specific genetic effects on these DMET genes from 1.5 million associations. To elucidate the phenotypic impact of sex-specific cis-eQTLs, we examined the sex-stratified genome-wide association studies (GWAS) in the UK Biobank. We found, for example, a SNP (rs34109652), associated with high cholesterol in male, also is a male-specific eQTL with *UGT2B17*, a gene that functionally annotated to metabolize testosterone. In contrast, these relationships were not observed in females. These findings allow us to construct a male-specific network of genetic regulation in *UGT2B17*, testosterone, and cholesterol. Lastly, we test sex heterogeneity of cis-DMET variants in 452 drug side-effect and human metabolic-related traits. We detected 40 traits that have at least one SNP shows sex heterogeneity. The most significant different loci are mapped to *ABCG2* associated with a greater risk to gout in males ($\hat{I}^2$-male = 0.028, $\hat{I}^2$-female = 0.0009, p-difference = 2.1e-183). Taken together, our work shows evidence of sex-specific variability in liver gene expression, genetic regulation, and phenotypic impact. Further elucidating the sex different interplay between genetic variants and gene expression of these DMET genes can enable precision medicine both in the identification of different disease etiology and in medication usage among sexes.

PrgmNr 1196 - Identifying disease-critical cell types and cellular processes across the human body by integration of single-cell profiles and human genetics

View session detail

Author Block: **A. L. Price**[1], K. Jagadeesh[2], K. K. Dey[1], D. T. Montoro[2], J. Engreitz[2], R. Xavier[2], A. Regev[2]; [1]Harvard Sch Pub Hlth, Boston, MA, [2]Broad Inst. of MIT & Harvard, Cambridge, MA
Disclosure Block: **A.L. Price:** None.

Cellular dysfunction is a hallmark of disease. Genome-wide association studies (GWAS) provide a powerful means to identify loci and genes contributing to disease risk, but in many cases the related cell types/states through which genes confer disease risk remain unknown. Deciphering such relationships is important both for our understanding of disease, and for developing therapeutic interventions.

Here, we introduce a framework for integrating single-cell RNA-seq (scRNA-seq), tissue-specific epigenomic maps and GWAS summary statistics to infer the underlying cell types and processes by which genetic variants influence disease. We analyzed 1.6 million scRNA-seq profiles from 209 individuals spanning 11 tissue types and 6 disease conditions, and constructed three types of gene programs: (i) cell type programs reflecting specific expression in the focal cell type; (ii) disease progression programs reflecting disease-specific differences in gene expression within the same cell type; and (iii) cellular process programs reflecting gene co-expression patterns within and across cell types. We evaluated these gene programs for relevance to disease by transforming them to SNP annotations with tissue-specific epigenomic maps and assessing disease heritability enrichment across 60 diseases and complex traits (average N=297K). We benchmarked our approach using blood cell traits, correctly recovering known cell type-trait pairs such as lymphocytes for WBC count (2.3x excess enrichment, $p = 3x10^{-5}$) and erythrocytes for RBC count (2.2x, $p = 2x10^{-7}$) while producing stronger signals than other approaches.

Results across a broader set of diseases and traits recapitulated known biology and produced novel findings. We highlight three examples. First, the GABAergic neuron cell type program was enriched for major depressive disorder (MDD) (4.0x, $p=1x10^{-4}$). GABAergic neurons regulate the brainâ□□s ability to control stress levels, which is the most prominent vulnerability factor in MDD. Second, the M cell disease progression program was enriched for ulcerative colitis (UC) (2.2x, $p=1.07x10^{-4}$). M cells surveil the lumen for pathogens and play a key role in immune-microbiome homeostasis. As a validation of this result, we determined that M cells expand dramatically in UC colon. Third, a disease-specific complement cascade cellular process program was enriched for multiple sclerosis (MS) (4.9x, $p=5.5x10^{-11}$), consistent with studies showing that complement activity is a marker for MS progression. In conclusion, our framework provides a powerful approach for identifying the cell types and cellular processes by which genetic variants influence disease.

PrgmNr 1240 - Interpreting Genome-Wide Association Studies of Inflammatory Bowel Disease Through the Lens of Single-Cell Sequencing

View session detail

Author Block: **M. Krzak**[1], D. L. Taylor[1], T. Alegbe[1], M. Ghouraba[1], M. Przybilla[1], R. McIntyre[1], T. Raine[2], C. Anderson[1]; [1]Wellcome Sanger Inst., Hinxton, United Kingdom, [2]Addenbrooke's NHS Hosptials, Cambridge, United Kingdom
Disclosure Block: **M. Krzak:** None.
Inflammatory bowel disease (IBD) is a complex disease characterised by chronic inflammation of the digestive tract. Genome-wide association studies (GWAS) have identified 241 risk loci significantly associated with the two common forms of IBD, Crohnâ□□s disease and ulcerative colitis. The vast majority of these risk loci reside in non-coding regions of the genome, and we only know which gene is dysregulated to increase risk of disease for a minority. This knowledge gap makes it difficult to draw insights into disease pathology and identify new candidate drug targets. To improve biological insights from IBD GWAS, we generated single cell RNA-sequencing data from ileal biopsies ascertained from 25 CD patients with active ileal inflammation and 26 non-IBD controls. We identified 49 different cell-types among the ~140K sequenced cells, including all major immune, enterocyte, secretory and mesenchymal populations (excluding granulocytes). Our optimized single-cell dissociation protocol reduced the effects of anoikis, enabling transcriptomes to be accurately profiled and placed along the entire crypt-villus axis, for the first time. We identified 797 unique genes differentially expressed between CD patients and controls, with notable expression differences in stem cell, secretory and enterocyte populations. Genes involved in interferon gamma signalling were enriched among those most frequently dysregulated across cell types. In an attempt to identify which of these expression differences are likely causal of disease, rather than simply a consequence of it, we integrated results from the latest IBD GWAS to assess the extent to which genes captured disease heritability, and in which cell-types. Genes specifically expressed in Tregs, monocytes and IL10-negative monocyte-derived macrophages captured a significant fraction of disease heritability, strongly implicating these cell-types in disease pathogenesis. We investigated which genes were driving these enrichment signals and identified candidate effector genes at many IBD risk loci, with follow-up analyses ongoing. Reassuringly, many confirmed IBD effector genes known to have a role in the normal functioning of these cell-types were found, including *NOD2, IL18RAP, IL23R, NCF4,* and *IL2RA*.

PrgmNr 1241 - Identifying genetic determinants of transcriptional response dynamics at single cell resolution

View session detail

Author Block: **J. Wei**[1], J. Resztak[1], P. Wu[2], E. Sendler[1], A. Alazizi[1], H. Mair-Meijers[1], R. Slatcher[3], S. Zilioli[4], X. Zhou[2], F. Luca[1], R. Pique-Regi[1]; [1]Ctr. for Molecular Med. and Genetics, Wayne State Univ., Detroit, MI, [2]Dept. of Biostatistics, Univ. of Michigan, Ann arbor, MI, [3]Dept. of Psychology, Univ. of Georgia, Athens, GA, [4]Dept. of Psychology, Wayne State Univ., Detroit, MI
Disclosure Block: **J. Wei:** None.
Single cell technologies enabled gene expression analysis at single cell resolution to study heterogeneous cell populations across conditions and individuals; yet few computational methods exist to profile response dynamics. Single cell data can capture more fine-grained details by analyzing trajectory dynamics, variance or other higher order moments which are crucial to gain a better understanding of the molecular underpinnings in inter-individual variation in drug response. Here we collected new data and developed a new supervised approach based on the diagonal linear discriminant analysis (DLDA), to construct a robust low dimensional representation of transcriptional response dynamics for each cell-type and treatment. The DLDA axis is then used to identify gene-by-environment interactions in the response dynamics. To evaluate the new method and to compare to other dynamic metrics, we activated peripheral blood mononuclear cells from 96 African American donors with phytohemagglutinin (PHA) or lipopolysaccharide (LPS), and treated with the glucocorticoid dexamethasone (DEX). We performed scRNA-seq for 292,394 cells and identified four major cell types: B-cell, Monocyte, NK-cell and T-cell. We employed negative binomial distribution to calculate RNA expression dispersion. We detected 1,409 genes with variable dispersion, most in monocytes. Effects on dispersion induced by PHA and LPS are negatively correlated with those induced by DEX, which implies that DEX suppresses the activated immune response through effects on both gene expression mean and dispersion. We discovered 504 genetic variants associated with gene expression dispersion, corresponding to 89 unique genes. Our new DLDA approach is able to capture gene expression dynamic changes in response to treatments, and identify different dynamic patterns for gene expression along the response pseudotime (DLDA axis). For example, in the T cells treated with DEX we identified four distinctive expression response patterns: low, decreasing, high and increasing, of which the first two are enriched in response to virus and the TNF pathways. To identify genetic variants that regulate the response to treatments we performed both response eQTL mapping and dynamic interaction eQTL mapping using the response pseudotime. We identified 79 and 59 Genes with response eQTLs for mean and dispersion and 841 genes with DLDA eQTLs, most in T cells. Our results shed light on the dynamics of gene expression in response to stimuli and across individuals. These dynamic processes fundamentally regulate the immune system and may contribute to inter-individual variation in immunological processes and diseases.

PrgmNr 1242 - Opposing evolutionary pressures drive clonal evolution and health outcomes in the aging blood system

View session detail

Author Block: **K. Skead**[1,2,3], A. Ang Houle[1], M. Fave[1,3], S. Abelson[1,2], M. Agbessi[1], V. Bruat[1], B. Lin[1], D. Soave[4,1], L. Shlush[5], S. Wright[6], J. Dick[7,2,1], Q. Morris[8], P. Awadalla[1,2]; [1]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [2]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [3]Vector Inst. for Artificial Intelligence, Toronto, ON, Canada, [4]Wilfrid Laurier Univ., Waterloo, ON, Canada, [5]Dept. of Immunology, Weizmann Inst. of Sci., Rehovot, Israel, [6]Dept. of Ecology and Evolutionary Biology, Univ. of Toronto, Toronto, ON, Canada, [7]Princess Margaret Cancer Ctr., Toronto, ON, Canada, [8]Computational and Systems Biology Program, Mem. Sloan Kettering Cancer Ctr., New York, NY
Disclosure Block: **K. Skead:** None.
The genetic architecture of blood is highly complex with germline polymorphisms, somatic point mutations, and larger chromosomal alterations playing a role in shaping the fitness of the immune cells. Many advances have been made in understanding how the age associated acquisition of point mutations and somatic structural variants (SSVs) in blood, termed Age Related Clonal Hematopoiesis (ARCH) predispose individuals to hematological cancer or cardiovascular disease. Yet, ARCH is commonly observed in healthy individuals and our ability to predict who is at risk of progressing to disease remains limited. Here, we interrogate the mutational and selective pressures within blood to better understand how the full spectrum of somatic changes within an individual impact clonal fitness and disease outcomes. We integrate deep learning and population genetics methods to evaluate the complex interplay of positive and negative selection in deeply sequenced blood samples from 92 individuals who subsequently progressed to blood cancer and 385 healthy controls. We are able to discriminate amongst evolutionary classes with high accuracy (86%) and capture signatures of purifying selection in the majority of individuals. The proportion of passenger to driver mutations appears to be critical in determining if the selective advantage conferred by a driver mutation is able to overwhelm purifying selection acting on passenger mutations thus allowing disease-predisposing clones to rise to dominance. Further, we evaluate how selection shapes the prevalence of large somatic structural variation in blood sampled from 10,000 individuals across over 20 different genetic ancestries. Using dense genotyping arrays, we capture SSVs among 20 populations and find that ARCH attributed to somatic structural variation is twice as frequent as previously reported with up to one in ten individuals harboring a large SSV. We estimate the rate at which SSVs accrue in blood cells and find that selection impacts the size and frequency of SSVs within individual blood populations. To determine the functional impact of clonal mutations on molecular phenotypes, we investigate the relationship between structural variation and the transcriptome. We show that gains, losses and copy number neutral variants impact gene expression distinctly, with stabilising selection shaping the penetrance of copy number alterations on gene expression. Through exploring a range of evolutionary models, our work shows how different classes of selection shape clonal dynamics in both healthy and pre-malignant blood thus enabling us to better understand why certain individuals are at a high risk of malignancy.

PrgmNr 1243 - Trans-ethnic meta-analysis and functional characterization highlights gene regulation underlying the genetic architecture of telomere length

View session detail

Author Block: **R. Keener**[1], M. Taub[2], M. Conomos[3], J. S. Weinstock[4], M. Arvanitis[1], C. W. Greider[5], R. Mathias[6], A. Battle[6], on behalf of the NHLBI TOPMed Consortium; [1]Johns Hopkins Univ., Baltimore, MD, [2]Johns Hopkins, Baltimore, MD, [3]Univ. of Washington, Seattle, WA, [4]Univ. of Michigan, Ann Arbor, MI, [5]Univ. of California Santa Cruz, Santa Cruz, CA, [6]Johns Hopkins Univ, Baltimore, MD
Disclosure Block: **R. Keener:** None.
Regulation of telomere length (TL) is critical for human health. Individuals with very short TL exhibit Short Telomere Syndromes, which present as organ failure of bone marrow, liver, or lungs, while individuals with very long TL are predisposed to cancer. Genome-wide association studies (GWAS) have yielded numerous TL-associated loci, but the mechanisms underlying most of these signals remain poorly characterized. Many are non-coding, and the genes and cell types mediating their effects have been unclear. We performed a trans-ethnic meta-analysis including TL GWAS results on a total of 215,273 individuals from European (1,4), Singaporean Chinese (2), South Asian (3,4), African (4), and Hispanic/Latino (4) ancestries and identified 56 genome-wide significant loci. To understand the mechanisms underlying these signals we performed fine-mapping, colocalization analysis with expression quantitative trait loci (eQTLs) from diverse tissues, and functional characterization. Using the European individuals from the meta-analysis, we conducted LD Score regression (LDSC) and identified blood/immune cells as key cell types for TL. One of our novel loci colocalizes with an eQTL for *TCL1A* in GTEx whole blood. TCL1A promotes cellular proliferation through the AKT pathway and overexpression of *TCL1A* is associated with T-cell leukemia and lymphoma. We detected a suggestive interaction between the putative causal SNP and age; stratified analysis by age groups demonstrated that the effect size for this SNP significantly increases with age. We have pursued functional characterization in human cell lines for several genes implicated by colocalization to examine their causal effect on telomere length. Among these, we have found evidence that proteins involved in RNA processing and cellular proliferation affect telomere length in cultured cells. Our results demonstrate mechanisms of novel genes associated with telomere length that will increase our understanding of human telomere length regulation mechanisms.1. Li et al. AJHG, 2020 2. Dorajoo et al. Nat Comm., 2019 3. Delgado et al. J Med Genet., 2018 4. Taub et al. bioRxiv, 2020

PrgmNr 1244 - Genetic associations from the COVID-19 Host Genetics Initiative highlight biology behind severe symptoms and infection susceptibility

View session detail

Author Block: **J. Karjalainen**, The COVID-19 Host Genetics Initiative; Inst. for Molecular Med. Finland, Helsinki, Finland
Disclosure Block: **J. Karjalainen:** None.
The ongoing SARS-CoV-2 pandemic is a global health concern. Knowledge of human biology underlying susceptibility and severity of COVID-19 post-infection is needed to inform development of therapeutics and better understand individual risk.
The COVID-19 Host Genetics Initiative (HGI) brings together the international human genetics community to generate, share and analyze data to identify the genetic determinants of COVID-19 susceptibility, severity, and outcomes. The initiative repeatedly performs meta-analysis of participating studies.
The HGI's analyses have replicated previous findings and implicated new biologically relevant variants for COVID-19 severity. For example, a TYK2 missense variant associated with protection from autoimmune disease and known to reduce TYK2 activity increases risk of COVID-19 related illness. Variants in the FOXP4 region, more common in non-European populations and showcasing the power of global analysis, that were previously associated with lung adenocarcinoma and ILD also confer risk to severe COVID-19 outcomes.
The HGI's sixth data freeze (spring 2021) consists of 61 studies from 24 countries, including many countries and ancestries typically underrepresented in genetic studies. The meta-analysis includes 9,376 cases of critical illness (respiratory support or death), 25,027 cases with severe symptoms (hospitalization), and 125,548 cases with lab-confirmed or self-reported PCR-confirmed infection.
With the current data we find additional genetic variation associated with severe COVID-19 symptoms. Curiously, rs35705950 at MUC5B, an strong risk variant for idiopathic pulmonary fibrosis, confers protection from severe symptoms (p 5.5e-9, OR 0.89). Other novel associations to severe symptoms include a lead missense variant (rs721917) in surfactant protein SFTPD previously associated to COPD (p 1.9e-8, OR 1.06), and a lead missense variant (rs117169628) in the lung-expressed transporter SLC22A31 which is co-expressed with surfactant protein genes (p=2.5e-8, OR 1.09).
Interestingly, in the current analysis of infected cases, we find a strong protective effect for rs190509934 69 bases from the transcriptional start site of ACE2, a receptor for the spike protein of SARS-CoV-2 (p 3.6e-18, OR 0.69, AF SAS 0.03, AF EUR 0.002), suggesting genetic variation at ACE2 is associated with protection from SARS-CoV-2 infection.
The initiative shows the power of quickly translating genetic data worldwide to biologically relevant findings for follow-up studies and creates an example for future global genetics efforts. The HGI's results are immediately made available at covid19hg.org with no restriction of use.

PrgmNr 1245 - Whole Genome Sequencing (WGS) based Association Study of 21 Inflammation Biomarkers in up to 38,473 Multi-Ethnic Individuals Identifies Novel Signals

View session detail

Author Block: **M. Jiang**[1], S. Gaynor[2], Y. Li[3], L. M. Raffield[4], P. L. Auer[5]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Harvard Univ., Boston, MA, [3]Univ North Carolina, Chapel Hill, NC, [4]UNC - Chapel Hill, Chapel Hill, NC, [5]Univ. of Wisconsin Milwaukee, Milwaukee, WI
Disclosure Block: **M. Jiang:** None.
Chronic inflammation is associated with a variety of diseases and health conditions including diabetes, cardiovascular disease, cancer, and allergic or autoimmune disorders such as asthma and rheumatoid arthritis. Genome-wide association studies (GWAS) have been performed on some inflammation biomarkers, such as c-reactive protein (CRP). But many others remain under-studied and mostly assessed in individuals of European ancestry. The latest CRP GWAS from Raffield et al. examined a multi-ancestry population (n = 23,279) with WGS data from the Trans-Omics for Precision Medicine (TOPMed) program and found evidence for eight distinct associations at the *CRP* locus. Here, we present a larger scale (up to n = 38,473 multi-ancestry samples) WGS-based association study for 21 inflammation biomarkers - CD40, CRP, E-Selectin, ICAM1, IL10, IL18, IL1, IL6, IL8, 8-Epi-PGF2 alpha, Lp-PLA2 activity, Lp-PLA2 mass, MCP1, MMP1, MMP9, MPO, OPG, P-Selectin, TNFa-r1, TNFa, and TNF-r2. We performed linear mixed model based single variant association analysis on variants with a minor allele count >= 10, using the GENESIS tool on the BioData Catalyst platform, adjusting for age, sex, race, study, and 11 genotype principal components. Using a genome-wide significance threshold of $1x10^{-9}$, we identified 44 unique loci, corresponding to 50 locus-trait pairs, for these 21 biomarker traits, ranging from zero locus for CD40, IL10, IL1, IL8, 8-Epi-PGF2 alpha, MMP1, MPO, OPG, TNFa-r1, and TNFa, to 16 loci for CRP, where a locus is loosely defined by genomic location. Conditioning on known variants for each trait if available, we identified the *MMP9* gene locus for MMP9. Surprisingly this association has not been previously reported. We also performed sequential conditional analysis to detect distinct signals at each identified locus, revealing multiple new distinct signals at known loci. Specifically, we detected 34 new distinct signals across 14 loci for 7 biomarkers - 4 new signals for CRP, 4 for P-Selectin, 4 for E-Selectin, 1 for IL6, 10 for ICAM1, 7 for Lp-PLA2 activity, and 4 for Lp-PLA2 mass. For example, for the phenotype P-Selectin, our analyses revealed 3 new distinct signals at the *SELP* locus, in addition to the 2 distinct signals previously reported by Barbalic et al. and Suhrefound et al. (2017); similarly at the *ABO* locus, our P-Selectin analyses identified one new signal distinct from the one previously reported. These results demonstrate the power of increased sample size and increased genetic diversity to identify new associations with inflammatory traits.

PrgmNr 1248 - Potential regulatory effect of the type 2 diabetes associated *KCNQ1* locus on *CDKN1C* expression identified using induced pluripotent stem cell derived pancreatic islet-like cells

View session detail

Author Block: **A. K. Nair**[1], Y. L. Muller[2], M. Traurig[2], J. Sutherland[1], R. G. Nelson[1], C. Bogardus[2], L. Baier[2]; [1]Natl. Inst. of Diabetes and Digestive and Kidney Diseases, NIH, Phoenix, AZ, [2]Natl. Inst. of Diabetes Digestive and Kidney Diseases, NIH, Phoenix, AZ
Disclosure Block: **A.K. Nair:** None.
American Indians from Southwest Arizona have a high incidence rate of type 2 diabetes (T2D). In GWAS for T2D in this population, the top T2D signal maps to intron 15 of *KCNQ1* which resides in a gene-rich region on chromosome 11. This region is highly imprinted (genes expressed solely from maternal or paternal allele) and several genes show tissue and developmental stage-specific imprinting. A previous study found that this T2D signal has a parent-of-origin effect in American Indians, with increased T2D risk and, among normal glucose tolerant individuals, a lower insulin secretion when the risk allele is inherited maternally. The genes in this region include *KCNQ1*, *CDKN1C*, *TRPM5*, *IGF2*, *TH*, *H19*, *INS* and others known to be important for islet development and function and may contribute to T2D. To identify the effector gene and causal SNPs at this locus, we obtained induced pluripotent stem cells (iPSC) derived from blood cells of American Indians who were clinically characterized and their DNA genotyped, and optimized a seven-stage differentiation protocol to generate glucose-responsive pancreatic islet-like cells. Using these iPSCs, we analyzed imprinting for *KCNQ1*, *CDKN1C*, *TRPM5* and *TH* during different stages of differentiation. *KCNQ1* had monoallelic expression during the iPSC stage (day 0), pancreatic progenitor stage (PP, day 10) and endocrine progenitor stage (day 13-16) while there was a gradual loss of imprinting thereafter. *CDKN1C* had monoallelic expression during all stages of differentiation. In contrast, *TRPM5* started losing imprinting during the PP stage and we observed bi-allelic expression during later stages while *TH* had biallelic expression during all stages. RNA sequencing data from different stages of pancreatic islet development identified 20 genes in a 1.2MB region around the T2D signal that are expressed during various stages of islet development. We then used CRISPR/CAS9 to generate isogenic iPSCs with hemizygous deletion in a region predicted to have regulatory function and has 4 SNPs, including the GWAS index SNP, by in-vitro assays and bio-informatic analysis. Differentiating this iPSC identified increased expression of *CDKN1C* during islet development compared to the parental cell line where the biggest difference (~2 fold) was seen on Day 13 while no difference was seen for the other genes, suggestive of a potential regulatory effect. However, we cannot rule of the possibility that the effect on *CDKN1C* expression may be independent of the SNPs. To further validate this, our ongoing work includes generation and differentiation of isogenic iPSCs with targeted edits at multiple SNPs in this region to assess the effect on *CDKN1C* expression.

PrgmNr 1249 - Single-cell gene expression and chromatin accessibility profiling of human pancreatic islets at basal and stimulatory conditions nominates mechanisms of type 1 diabetes genetic risk

View session detail

Author Block: **R. D'Oliveira Albanus**[1], X. Tang[2], H. Taylor[3], M. Erdos[3], N. Narisu[3], F. S. Collins[3], S. Chen[2], S. Parker[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Weil Cornell Med., New York, NY, [3]NIH, Bethesda, MD

Disclosure Block: **R. D'Oliveira Albanus:** None.

Type 1 diabetes (T1D) is an autoimmune disease arising from a complex interplay of genetics and environmental factors. The role of beta cells in the genetics of T1D predisposition is not yet fully understood. Here, we investigated how genetics and the pancreatic islet chromatin and transcriptome landscapes mediate T1D risk. We performed single-cell multi-omics (chromatin accessibility and gene expression) profiling of healthy (n=8), pre-T1D (n=2), and T1D (n=1) human islets. Healthy islets (n=3) were additionally profiled under cytokine stimulation (TNF-Î± and IL-1Î²) and CVB4 infection to mimic T1D environmental effects. Joint integration of all datasets (n=34) yielded ~121k cells (~50k snATAC, ~71k scRNA). We identified 10 different cell types, ranging from 1.4% (immune) to 35% (ductal) of all cells. Beta cells (13% of all cells) expression analyses identified sets of differentially expressed genes (DEGs) across healthy/disease/environmental conditions. Environmental DEGs shared pathway-level enrichments with T1D DEGs, indicating molecular concordance between *in vivo* disease state and *in vitro* disease models. Comparing T1D GWAS summary statistics to basal chromatin profiles identified immune cells with the highest enrichment (log enrichment = 2.78; fGWAS), and also identified enrichment in beta, alpha, acinar, ductal, and quiescent stellate cells (log enrichment range 1.53 to 2.12; fGWAS). Differentially accessible regions in pre-T1D beta cells were more enriched for T1D GWAS signals compared to non-differentially accessible regions in the same cell type (log enrichment = 3.43 and 1.75, respectively; fGWAS). A similar trend was observed in alpha and delta cells, indicating a role for non-immune islet cells mediating T1D risk at early disease stages. Functional fine-mapping using chromatin accessibility prioritized three T1D GWAS variants at the *DLK1/MEG3*, *TOX*, and *RASGRP1* loci overlapping beta- or islet-specific regulatory elements. Beta cell allelic effects in these loci were additionally supported by a predictive model, which was validated with allele-specific chromatin accessibility analyses. Condition-specific co-accessibility analyses nominated four effector transcripts at these three loci (gene-SNP distance 0-159 kb). We are currently generating human induced pluripotent stem cells knockout lines and differentiating them into beta cells to functionally validate the nominated genes and regulatory elements. We plan to present these validation results during the meeting. Together, our work reveals the environmental context specificity of T1D genetic risk mediated through endocrine islet cells.

PrgmNr 1250 - The Kidney genome atlas reveals a novel locus on chromosome 14 associated with adult proteinuric kidney diseases

View session detail

Author Block: **E. Fast**[1], T. W. Soare[1,2], H. Somineni[1,2], J. Watkins[1], A. Tebbe[1], M. Lalioti[1], L. Walsh[1], M. Kretzler[3], G. Nadkarni[4], R. Gbadegesin[5], J. Wenke[6], D. G. MacArthur[7], P. Mundel[1,8], T. T. Tibbitts[9,1], M. Penny[1]; [1]Goldfinch Bio, Cambridge, MA, [2]Insitro, South San Francisco, CA, [3]Univ. of Michigan, Ann Arbor, MI, [4]Mount Sinai Hosp., New York City, NY, [5]Duke Univ. Med. Ctr., Durham, NC, [6]Nashville BioSci.s, Nashville, TN, [7]Garvan Inst., Darlinghurst, Australia, [8]Atlas Venture, Cambridge, MA, [9]Dogodan Therapeutics, Inc (DTI), WESTFORD, MA

Disclosure Block: **E. Fast:** Salary/Employment; Goldfinch Bio.

Chronic Kidney disease (CKD) affects 1 in 9 people worldwide. There is a high unmet need for drugs that extend and restore kidney function, because dialysis and organ transplantation carry substantial economic and psychological burden. To foster drug development of genetically validated targets, we have created the Kidney Genome Atlas (KGA) by assembling ~23,000 whole genomes from 2,832 kidney disease cases including proteinuric kidney disease cases such as Focal segmental glomerulosclerosis (571 cases), minimal change disease (244 cases), nephrotic syndrome (196 cases) and idiopathic proteinuria (1,123 cases) and 19,804 controls. Following the gnomAD pipeline, we implemented a rigorous quality control procedure to obtain a high confidence dataset for downstream analyses of proteinuric kidney diseases. Ancestries were inferred genetically based on a k-NN model trained on 1,000 Genomes data which resulted in 597 cases and 10,127 controls of European (EUR) ancestry, 513 cases and 3,805 controls of African (AFR) ancestry, and 290 cases and 754 controls of Latino/Admixed American (AMR) ancestry for association testing. Meta-analysis of common variants across ancestries showed minimal impact of potential confounders, such as ancestry or sequencing center differences (lambda=1.03). We identified a novel locus on chromosome 14 (rs11160484; effect size = -0.42, P = $2.8*10^{-8}$) associated with proteinuric kidney disease. In addition, we confirmed the well-known association of APOL1 risk haplotypes (G1/G1, G2/G2 or G1/G2; effect size = 0.50, P = $2.4*10^{-9}$, under recessive model) in the AFR cohort. LD-score regression analysis revealed a trend towards a weak positive genetic correlation (rg = 0.097, 90% CI [0.010, 0.18]) between proteinuric kidney diseases and CKD defined by estimated glomerular filtration rate or eGFR (Wuttke et al, 2019). Using summary statistics from our EUR dataset, we estimated the SNP heritability of proteinuric kidney diseases at 0.15 (95% CI [0.095, 0.20]), suggesting that there may be many more genetic contributions that are yet to be discovered. These findings advance our understanding of the genetic architecture of proteinuric kidney diseases and highlight an opportunity for novel therapies and patient stratification.

PrgmNr 1251 - Computational and functional gene prioritization from a saturated GWAS of adult height in >5 million people

View session detail

Author Block: **J. N. Hirschhorn**[1], N. Renthal[1], L. Yengo[2], S. Vedantam[3], E. Marouli[4], J. Baronas[5], P. Nakka[6], A. Eliasen[7], E. Bartell[8], S. Sakaue[9], T. Frayling[10], H. Kronenberg[11], G. Lettre[12], Y. Okada[13], A. R. Wood[14], GIANT consortium; [1]Boston Children's Hosp./Broad Inst., Boston, MA, [2]Brisbane, Australia, [3]Boston Children's Hosp, Sharon, MA, [4]Barts and The London Sch. of Med. and Dentistry Queen Mary Univ. of London, London, United Kingdom, [5]Boston Children's Hosp., Boston, MA, [6]Boston Childrens Hosp., Boston, MA, [7]Denmark Technical Univ., Copenhagen, Denmark, [8]Brookline, MA, [9]Osaka Univ., Suita, Japan, [10]Univ. of Exeter, Exeter, Devon, United Kingdom, [11]Massachusetts Gen. Hosp., Boston, MA, [12]montreal Heart Inst, Montreal, QC, Canada, [13]Osaka Univ. Graduate Sch. of Med., Suita, Osaka, Japan, [14]Coll. of Med. and Hlth., Univ. of Exeter, Exeter, Devon, United Kingdom
Disclosure Block: **J.N. Hirschhorn:** Consultant/Consulting Fees/Other Remuneration; Camp4 Therapeutics.

Adult height, the classical polygenic trait, is the result of childhood skeletal growth. A prior abstract described the GIANT consortium GWAS of height, now up to >5 million individuals including >1 million of non-European ancestry, that has identified >12,000 approximately independent signals of association. In European-ancestry individuals, these signals account for ~40% of variation in height, ~80% of the heritability attributable to common variation (h^2_{SNP}). Across multiple ancestries, >90% of h^2_{SNP} maps to the ~23% of the genome that lies within 35 kb of these signals. We have now taken computational and experimental approaches to prioritize genes likely underlying the height associations. We used DEPICT, PoPS, and MAGMA, separately and in combination, to prioritize genes from height GWAS. The combination of all 3 methods yielded a set of 356 genes; these scored better than genes prioritized by fewer methods, as assessed by multiple metrics. These include per-SNP heritability (LD score regression normalized Ï=4.38, 95% CI=2.35-6.41); in addition, 54% of the 356 genes were nearest to a GWAS signal and 16% were âOMIM genesâ that underlie monogenic skeletal growth disorders. We also analyzed expression data relevant to growth plate chondrocytes, a key cell type for skeletal growth. First, we compared gene expression for microdissected samples from the 3 main layers of the mouse growth plate: round, flat, and hypertrophic. For each gene, we computed specificity for each layer as (layer expression/total expression) and also MAGMA gene-based *p* values from height GWAS. Specificity in all 3 layers was significantly correlated with GWAS *p* values; this was driven by genes with specificity for the least differentiated layer (round cells, *p*-6 conditioned on other layers). We also generated expression data from an *in vitro* murine model of chondrocyte differentiation and showed that specificity for expression earlier in differentiation was correlated with GWAS *p* values (*p*-5, conditioned on late expression). Finally, we used this *in vitro* system to perform a pooled CRISPR screen to identify genes that, when perturbed, accelerate or delay differentiation (assayed by cell surface CD200, which appears late in differentiation). The 10 genes that most strongly accelerate maturation when perturbed include OMIM genes (*EED*, *EZH2*, *SUZ12*), known regulators of growth plate maturation (*PTCH1*, *SUFU*), and genes prioritized by all 3 computational methods (*SUFU*, *DOT1L*, *SMAD6*). In summary, a combination of GWAS, experimental data and computational approaches identifies genes with multiple lines of evidence for regulating skeletal biology and height.

PrgmNr 1252 - Genetic and functional characterisation of the Melanocortin 3 Receptor (*MC3R*) links nutritional state to growth, body composition and the onset of puberty

View session detail

Author Block: **A. Williamson**[1,2], B. Y. H. Lam[1], S. Finer[3], F. R. Day[2], J. A. Tadross[1,4], A. GonÃ§alves Soares[5], K. H. Wade[5], A. Melvin[1], G. K. C. Dowsett[1], K. E. Ridley[6], Z. Xu[6], I. Cimino[1], D. Rimmington[1], K. Rainbow[1], K. Duckett[1], S. Holmqvist[6], A. Khan[3], Genes & Health Research Team, R. C. Trembath[7], H. C. Martin[8], A. P. Coll[1], D. H. Rowitch[6], N. J. Wareham[2], D. A. van Heel[3], N. J. Timpson[5], K. K. Ong[2,6], C. Langenberg[2,9], J. R. B. Perry[2], G. S. Yeo[1], S. O'Rahilly[1]; [1]MRC Metabolic Disease Unit, Wellcome-MRC Inst. of Metabolic Sci., Univ. of Cambridge, and NIHR Cambridge BioMed. Res. Ctr., Cambridge, United Kingdom, [2]MRC Epidemiology Unit, Wellcome-MRC Inst. of Metabolic Sci., Univ. of Cambridge, Cambridge, United Kingdom, [3]Inst. of Population Hlth.Sci., Barts and the London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom, [4]Dept. of Pathology, Univ. of Cambridge, Cambridge, United Kingdom, [5]MRC Integrative Epidemiology Unit and Population Hlth.Sci., Bristol Med. Sch., Univ. of Bristol, Bristol, United Kingdom, [6]Dept. of Paediatrics, Univ. of Cambridge, Cambridge, United Kingdom, [7]Sch. of Basic and Med. BioSci.s, Faculty of Life Sci. and Med., Kingâ□□s Coll. London, London, United Kingdom, [8]Wellcome Sanger Inst., Cambridge, United Kingdom, [9]Computational Med., Berlin Inst. of Hlth., CharitÃ© Univ. Med. Berlin, Berlin, Germany
Disclosure Block: **A. Williamson:** None.

The state of somatic energy stores in metazoan organisms is communicated to the brain, which regulates key aspects of behaviour, growth, nutrient partitioning and development. The central melanocortin system is an important component of this, with the Melanocortin 4 Receptor (MC4R) controlling appetite, food intake and energy expenditure. Humans lacking *MC4R* are obese, hyperphagic and have reduced basal energy expenditure. Melanocortin 3 Receptor (MC3R) is the only other melanocortin receptor that is predominantly expressed in the brain. Rare functionally compromised heterozygous variants in *MC3R* have been reported in humans, however no clear consistent phenotype has been reported.

We aimed to establish the role of MC3R in human physiology using a reverse genetics approach. We identified naturally occurring coding mutations in *MC3R* in 3 population-based cohorts - Avon Longitudinal Study of Parents and Children (ALSPAC), UK Biobank and Genes & Health using exome sequencing and array-based genotyping. We functionally studied these mutations *in vitro* to identify those resulting in functional impairment of the receptor and subsequently studied the relationship of these variants with human phenotypes. We further characterised the population of *Mc3r* expressing neurons in the mouse hypothalamus using single cell RNA-sequencing (scRNA-seq) and RNA-scope.

We found that humans who carry characterised loss of function mutations in MC3R, including a single homozygote consanguineous individual, have a later onset of puberty. These individuals also had reduced linear growth, having reduced height in both childhood and adulthood. The carriers of these mutations had reduced lean mass and lower circulating IGF-1 levels, an indirect measure of growth hormone.

scRNA-seq of the mouse hypothalamus and RNA-scope show that within the hypothalamus *Mc3r* expression is enriched in sub-populations of neurons expressing either Growth hormone-releasing hormone (*Ghrh*) or Kisspeptin/Tachykinin B (*Kiss1*/*Tac2*) which regulate growth and reproductive function, respectively.

Across the animal kingdom, nutritional status is a critical determinant of linear growth and the timing of reproductive maturity. MC3R appears to play an important role in linking signals of caloric sufficiency that act through POMC expressing neurons to the central control of growth and reproduction.

PrgmNr 1253 - Causal signatures in skeletal muscle single nucleus multi-omics data across 287 individuals

View session detail

Author Block: **A. Varshney**[1], N. Manickam[1], M. Erdos[2], P. Orchard[1], A. U. Jackson[1], J. Okamoto[1], S. Hanks[1], D. Ciotlos[1], H. Stringham[1], N. Narisu[3], B. Lori[4], J. Welch[1], J. Chen[5], I. Barroso[6], C. P. Jenkinson[7], L. Norton[7], K. Williams[8], M. Laakso[9], J. Tuomilehto[10], T. A. Lakka[9], H. A. Koistinen[10], K. L. Mohlke[11], M. Boehnke[1], F. S. Collins[4], L. J. Scott[12], S. Parker[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]NHGRI (NIH), Bethesda, MD, [3]NIH, Bethesda, MD, [4]Natl. Human Genome Res. Inst., Bethesda, MD, [5]Wellcome Sanger Inst., Cambridge, United Kingdom, [6]RILD Building, Exeter, United Kingdom, [7]UTHSCSA, San Antonio, TX, [8]Henry Ford Hlth.System, Detroit, MI, [9]King Abdulaziz Univ., Jeddah, Saudi Arabia, [10]Univ. of Helsinki, Helsinki, Finland, [11]Univ North Carolina, Chapel Hill, NC, [12]Univ Michigan, Ann Arbor, MI
Disclosure Block: **A. Varshney:** None.
Skeletal muscle is the largest organ, by weight (>40%) and relevant for several polygenic traits including insulin resistance and type 2 diabetes (T2D). Identifying genetic mechanisms orchestrating these traits requires pinpointing the set of causal variants, the target genes these variants regulate, and the specific cell populations in which they function. Here, we generated single nucleus chromatin (snATAC) and transcription (snRNA) profiles across 287 frozen human skeletal muscle biopsies. We integrated 414,982 nuclei across the modalities with LIGER and identified 13 cell types, ranging in abundance from 38.1% (type 1 fiber) to 0.5). Using summary-based mendelian randomization (SMR) techniques, we inferred causality between chromatin accessibility and gene expression. We identified 6,190 caQTL peaks (caPeaks) that were causal on 1,299 eQTL genes (eGenes), and 1,277 eGenes that were causal on 6,236 caPeaks (SMR 5% FDR, pHEIDI > 0.05). caPeaks causal on eGenes are further from the gene TSS than target caPeaks (Wilcoxon rank sum test P=9e-4), suggesting that distal caPeaks are enhancers that causally affect gene expression whereas proximal caPeaks can be modulated by the transcription machinery near gene TSSs. We performed colocalization and chain-of-causality inferences of e/caQTL with various genome-wide association study (GWAS) signals across the five cell types. Of 100 T2D GWAS signals that colocalize with caQTL, 48 are cell type-specific, highlighting the importance of sn-caQTL maps for GWAS functional studies. We identified 287 caPeaks that mediated effects on 129 eGenes which then mediated effects on T2D. Studying COVID19 GWAS, we observed that the IFNAR2 signal colocalized with a caQTL in the region, and eQTLs for both IFNAR2 and the neighboring IL10RB cytokine receptor genes. The caQTLs are causal on both the eQTLs that are causal on COVID19. These results inform the causal chain of molecular events that influence the complex genetic regulatory architecture of skeletal muscle at high-resolution epigenomic, transcriptomic, and cell state scales.

PrgmNr 1256 - A global biobank study of asthma identifies novel associations, illuminates shared genetic architecture, and improves polygenic prediction across diverse ancestry groups

View session detail

Author Block: **K. Tsuo**[1,2,3,4], W. Zhou[3,2,4], Y. Wang[3,2,4], M. Kanai[4,3], B. Neale[3,2,4], M. Daly[1,2,3,4,5], A. R. Martin[3,2,4], Global Biobank Meta-analysis Initiative; [1]Dept. of Genetics, Harvard Med. Sch., Boston, MA, [2]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp. Res. Inst., Boston, MA, [4]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, [5]Inst. for Molecular Med. Finland, Helsinki Inst. of Life Sci., Univ. of Helsinki, Helsinki, Finland
Disclosure Block: **K. Tsuo:** None.
Asthma is a complex and multifactorial disease that affects millions of people worldwide and varies in prevalence by an order of magnitude across geographic regions and diverse populations. However, the extent to which genetic variation contributes to these disparities is unclear, as studies probing the genetics of asthma have been primarily limited to populations of European descent. To expand our understanding of the genetic factors underlying asthma risk in different ancestral populations, we conducted the largest genome-wide association study of asthma to date (N cases=153,763 and N controls=1,647,022) via meta-analysis across 18 biobanks with harmonized phenotype definitions and spanning multiple countries and genetic ancestries, collectively called the Global Biobank Meta-analysis Initiative (GBMI). This meta-analysis discovered 180 independent genome-wide significant loci (p

PrgmNr 1257 - Genome-wide polygenic risk score of prostate cancer in African and European ancestry men

View session detail

Author Block: **B. F. Darst**[1], R. K. Madduri[2], A. Rodriguez[2], X. Sheng[1], R. Eeles[3], Z. Kote-Jarai[3], J. M. Gaziano[4], A. C. Justice[5], D. V. Conti[1], C. A. Haiman[1], Million Veterans Program, PRACTICAL Consortium; [1]Univ. of Southern California, Los Angeles, CA, [2]Univ. of Chicago, Chicago, IL, [3]The Inst. of Cancer Res., London, United Kingdom, [4]VA Boston Hlth.care System, Roxbury Crossing, MA, [5]VA Connecticut Hlth.care System, West Haven, CT

Disclosure Block: **B.F. Darst:** None.

Genome-wide polygenic risk scores (PRS) are reported to have higher performance than standard genome-wide significant PRS across numerous traits. We evaluated the ability of genome-wide PRS to evaluate prostate cancer risk compared to our recently developed and highly predictive multi-ancestry PRS of 269 established prostate cancer risk variants. Genome-wide PRS approaches included LDpred2, PRS-CSx, and EB-PRS. Models were trained using the largest and most diverse prostate cancer GWAS to date of 107,247 cases and 127,006 controls, which was previously used to develop the multi-ancestry PRS of 269 variants. Resulting models were tested in independent samples of 1,586 cases and 1,047 controls of African ancestry from the California Uganda Study and 8,045 cases and 191,835 controls of European ancestry from the UK Biobank. Among the genome-wide PRS approaches, LDpred2 had the best performance, with AUCs of 0.649 (95% CI=0.627-0.670) in African and 0.819 (95% CI=0.815-0.823) in European ancestry men. African and European ancestry men in the top PRS decile relative to men in the median 40-60% PRS category had odds of prostate cancer of 3.29 (95% CI=2.47-4.40) and 2.99 (95% CI=2.78-3.23), respectively. However, the PRS constructed using 269 variants had significantly larger AUCs in both African (0.679, 95% CI=0.659-0.700) and European ancestry men (0.845, 95% CI=0.841-0.849), with African and European ancestry men in the top PRS decile having larger odds of prostate cancer (3.53, 95% CI=2.66-4.69 and 4.20, 95% CI=3.89-4.53, respectively). We are currently further validating these findings in diverse men from Million Veteranâ□□s Program. This investigation suggests that genome-wide PRS may not improve the ability to distinguish prostate cancer compared to a genome-wide significant PRS.

PrgmNr 1258 - Genetic association of phenotypes derived by self-supervised deep learning of retina fundus images reveals new genes for eye development

View session detail

Author Block: **Z. Xie**[1], T. Zhang[1], S. Kim[1], W. Zhang[2], C-H. Lin[3], M-R. Wu[3], A. Davis[3], L. Giancardo[2], H. Chen[2], S. Wang[3], R. Chen[1], D. Zhi[2]; [1]Baylor Coll. of Med., Houston, TX, [2]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [3]Stanford Univ., Stanford, CA
Disclosure Block: **Z. Xie:** None.
Although genome-wide association studies (GWAS) have achieved great success and identified thousands of genetic associations, phenotypes of most existing GWAS studies are predefined. While these phenotypes encode valuable biomedical knowledge, they are also biased by current clinical practice and epidemiological studies. Also as phenotype code is greatly simplified, it is often not sufficient to capture the complexity of human physiology and pathology in their entirety. Fortunately, with the medical record becoming increasingly digitized, there are new opportunities to derive phenotypes beyond expert-curation, which would avoid human bias and discover new phenotypes that are previously missed. Here, leveraging breakthroughs in self-supervised deep representation learning, we propose a new approach for phenotype discovery from medical images. We use a contrastive loss function over an Inception V3 architecture to learn a representation that captures the inherent image features of individuals. Using vessel segmentation masks generated from retina fundus images as inputs, we designed a phenotyper neural network model that generates 128 phenotypes representing retinal vasculature. After training on 40,000 images from EyePACS, our model generated phenotypes from 130,967 images of 65,629 British White participants in the UK Biobank. A GWAS of these vasculature phenotypes identified 34 independent loci, at least 5 are associated with vessel features. Mouse knockout experiments verified the role of the WNT7B gene, a newly found locus, in retinal vessel development. Our results establish a new framework of unsupervised image based genome wide genotype phenotype association studies (iGWAS). Our framework would expand the repertoire of GWAS phenotypes and enable discovery of new biology.

PrgmNr 1259 - Fine-mapping across diverse ancestries drives the discovery of putative causal variants underlying human complex traits and diseases

View session detail

Author Block: **K. Yuan**[1,2,3], T-T. Chen[4], S-C. Lin[4], R. Longchamps[1,2,3], A. F. PardiÃ±as[5], M. Lam[1,3,6,7,8], C-Y. Chen[9], Y-C. Feng[1,2,10,11], Y-F. Lin[4,12,13], T. Ge[2,10,14], H. Huang[1,2,3]; [1]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [2]Stanley Ctr. for Psychiatric Res., the Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Dept. of Med., Harvard Med. Sch., Boston, MA, [4]Ctr. for Neuropsychiatric Res., Natl. Hlth.Res. Inst.s, Miaoli, Taiwan, [5]MRC Ctr. for Neuropsychiatric Genetics and Genomics, Cardiff Univ. Sch. of Med., Cardiff, United Kingdom, [6]Human Genetics, Genome Inst. of Singapore, Singapore, Singapore, [7]Div. of Psychiatry Res., the Zucker Hillside Hosp., Northwell Hlth., Glen Oaks, NY, [8]Res. Div. Inst. of Mental Hlth.Singapore, Singapore, Singapore, [9]Biogen, Cambridge, MA, [10]Psychiatric and Neurodevelopmental Genetics Unit, Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [11]Vertex Pharmaceuticals, Boston, MA, [12]Dept. of Publ. Hlth.& Med. Humanities, Sch. of Med., Natl. Yang Ming Chiao Tung Univ., Taipei, Taiwan, [13]Inst. of Behavioral Med., Coll. of Med., Natl. Cheng Kung Univ., Tainan, Taiwan, [14]Dept. of Psychiatry, Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA

Disclosure Block: **K. Yuan:** None.

Genome-wide association studies (GWAS) of human complex traits or diseases often implicate genetic loci that span hundreds of significant genetic variants. However, these loci may only contain one or a handful of causal variants. Statistical fine-mapping reï¬nes a GWAS locus to a smaller set of likely causal variants (credible set). Since non-causal variants have marginally different effects across populations where LD differs, capitalizing on the genomic diversity across ancestries holds the promise to further improve the resolution of fine-mapping. However, to date, cross-population fine-mapping efforts have been limited, partly due to the lack of statistical methods that can appropriately integrate data from multiple ancestries. Building on Sum of Single Effects (SuSiE), a single-population fine-mapping model, we have developed SuSiEx, an accurate and computationally efficient method for trans-ancestry fine-mapping. SuSiEx assumes that causal variants are largely shared across populations, while allowing for varying variant effect sizes across populations. Our model can integrate data from an arbitrary number of ancestries, explicitly models population-specific LD patterns, accounts for multiple causal variants in a genomic region, and can be applied to GWAS summary statistics without access to individual-level data. We showed, via simulation studies, that compared with fine-mapping 100K European samples, integrating 50K European and 50K African samples using SuSiEx enabled fine-mapping of more association signals, and dramatically increased the resolution of credible sets. Comparing with PAINTOR, SuSiEx had a 37% reduction in the median size of credible sets and a 54% increase in the number of high Posterior Inclusion Probability (PIP) variants. We applied SuSiEx to 25 quantitative traits that are available from both the Taiwan Biobank (TWB, n = 92,615) and UK Biobank (UKBB, n = 361,194) to fine-map genetic loci reaching genome-wide significance. Compared with single-population fine-mapping in UKBB, cross-ancestry fine-mapping significantly reduced the size of credible sets and increased the PIP of the most probable variant. We additionally applied our method to schizophrenia GWAS summary statistics of East Asian and European ancestries. Compared with the published fine-mapping results from PGC using FINEMAP on the same data, SuSiEx reduced the size of credible sets in 70% of the fine-mapped loci. Manual inspection confirmed that SuSiEx provided more sensible results in many loci. As the accumulation of GWAS results from different ancestries, the application of our method will be much promising.

PrgmNr 1260 - Analysis across Taiwan Biobank, Biobank Japan and UK Biobank identifies hundreds of novel loci for 36 quantitative traits

View session detail

Author Block: **Y-F. Lin**[1], C-Y. Chen[2], T-T. Chen[1], Y-C. Feng[3], R. Longchamps[4], S-C. Lin[1], M. J. Daly[5], W. J. Chen[6], H. Huang[4], T. Ge[7]; [1]Natl. Hlth.Res. Inst.s, Zhunan Town, Miaoli County, Taiwan, [2]Biogen, Cambridge, MA, [3]MGH; Broad Inst., Cambridge, MA, [4]Boston, MA, [5]Massachusetts Gen. Hosp., Boston, MA, [6]Natl. Taiwan Univ, Taipei, Taiwan, [7]Harvard Med. Sch., Malden, MA
Disclosure Block: **Y. Lin:** None.
Genome-wide association studies (GWAS) have identified tens of thousands of genetic loci associated with human complex traits and diseases. However, the majority of GWAS were conducted in individuals of European (EUR) ancestry. Failure to capture global genetic diversity has limited biological discovery and impeded equitable delivery of genomic knowledge to diverse populations. Here we performed genome-wide analysis on 102,900 individuals across 36 human quantitative traits in the Taiwan Biobank (TWB), a major biobank effort that broadens the population diversity of genetic studies in East Asia (EAS). We identified 1,907 independent genome-wide significant loci (P-value -8) across the 36 traits, among which 1,287 loci survived Bonferroni correction for the number of traits tested (P-value -8/36). The number of genome-wide significant loci per trait ranged from 1 for forced expiratory volume in one second (FEV1) and FEV1 to forced vital capacity (FVC) ratio (FEV1R), to 211 for height (HT). We estimated the SNP-based heritability (h^2g) for each trait, which ranged from 0.009 (FEV1R) to 0.384 (HT), and pairwise genetic correlations between these traits, which identified clusters of highly genetically correlated traits. Of the 1,907 genome-wide significant loci, 1,615 were fine-mapped to a total of 1,972 credible sets, each representing an independent association signal. Out of the 1,972 credible sets, 232 were mapped to a single variant with posterior inclusion probability (PIP) > 95%, among which 24 were missense variants. Leveraging GWAS summary statistics from Biobank Japan (BBJ) and UK Biobank (UKBB), we found that the genetic architecture of the quantitative traits examined was largely consistent within EAS and between EAS and EUR populations. Integrating TWB and BBJ GWAS identified a total of 2,975 genetic loci, among which 979 had not been reported in previous biobank studies. We also examined whether polygenic risk scores (PRS) of biomarkers can be used to predict the risk of common complex disease, and demonstrated the potential utility of biomarker GWAS in predicting disease risk (e.g., type 2 diabetes) and the promise of multi-trait cross-population polygenic prediction. Our novel findings represent a major advance in diversifying GWAS samples and the characterization of the genetic architecture of human complex traits in EAS populations. Future endeavors on increasing the sample size and phenotype coverage in TWB, and improving cross-biobank data harmonization will further facilitate genomic discovery.

PrgmNr 1261 - The effects of demographic-based selection bias on GWAS results in the UK Biobank

View session detail

Author Block: **S. van Alten**[1,2], B. Domingue[3], T. Galama[1,2,4,5], A. Marees[1]; [1]Vrije Univ.it Amsterdam, Amsterdam, Netherlands, [2]Tinbergen Inst., Amsterdam, Netherlands, [3]Stanford Univ., Stanford, CA, [4]Erasmus Univ., Rotterdam, Netherlands, [5]Univ. of Southern California, Dornsife, Los Angeles, CA
Disclosure Block: **S. van Alten:** None.
Genome-wide association studies (GWASs) are almost always based on a non-random sample of the underlying population, as obtaining very large sample sizes, rather than ensuring such samples are representative, has been key to their success. Selection bias in estimated genetic associations, including how it varies across traits, is poorly understood. A sample of particular interest is the widely used UK Biobank (UKB). Because of the need for very large samples, the UKB is included in almost all large GWASs as one of the largest cohorts. In addition, UKB's subsample of genotyped siblings (UKBSIB) has become a crucial resource for estimating genetic effects free of environmental confounding. Using nationally representative UK Census microdata as a reference, we document substantial non-random selection into the UKB, and even stronger for UKBSIB: individuals in the UKB and UKBSIB are more likely to be female, higher educated, and older, compared to the underlying population that received an invitation. We also show that this non-random selection leads to significant selection bias in associations between various demographic and health-related traits estimated in the UKB. We then estimate probabilities of UKB participation for each UKB participant to estimate selection-corrected GWASs for multiple traits using inverse probability weighting. Based on preliminary analyses for the top 5,000 SNPs associated with BMI, education, and height, respectively, we show that the extent to which selection-corrected GWAS results differ from those of regular GWASs is trait-specific. Genetic associations for educational attainment and BMI are the most altered after correcting for volunteer bias, whereas associations for height remain relatively unaffected. For educational attainment, 12.6% of our estimated SNP effects flip sign after correcting for selection bias, suggesting that current GWAS methods are not sufficiently robust. We will extend these analyses by investigating more phenotypes, conducting regular and inverse probability weighted GWASs in the UKB that incorporate all available SNPs, and comparing results. Our findings will be useful for understanding the extent to which a particular phenotype is prone to selection bias in GWAS, and our correction method provides an alternative when population-representative cohorts are not available.

PrgmNr 1264 - *HDAC9* structural variants disrupting *TWIST1* transcriptional regulation lead to craniofacial and limb malformations

View session detail

Author Block: **R. Y. Birnbaum**[1], N. Hirsch[1], I. Dahan[1], E. D'haene[2], M. Avni[1], S. Vergult[3], P. Magini[4], C. Graziano[5], G. Severi[5], E. Bonora[6], F. Brancati[7], A. FernÃindez-JaÃ©n[8], O. J. Rory[9], B. Hallgrimsson[10]; [1]Ben Gurion Univ., Beer Sheva, Israel, [2]Ctr. for Med. Genetics, Ghent Univ., Ghent, Belgium, [34]Genetica Medica, Univ. of Bologna, Bologna, Italy, [5]U.O. Genetica Medica, IRCCS Azienda Ospedaliero-Univ.ria di Bologna, Bolonga, Italy, [6]Policlinico S Orsola, Bologna, Italy, [7]CSS - Mendel Inst., Rome, Italy, [8]Ben Gurion Univ., Hospital Universitario QuirÃ³nsalud, School of Medi, Spain, [9]Ctr. for Individualized Med., Mayo Clinic, Rochester, MN, [10]Univ Calgary, Calgary, AB, Canada

Disclosure Block: **R.Y. Birnbaum:** None.

Structural variants (SVs) such as insertions, deletions duplications, translocations and inversions, are associated with human disorders. SVs can affect protein coding sequences as well as gene regulatory elements. However, SVs disrupting protein coding sequences that also function as cis regulatory elements remain largely uncharacterized. Here, we show that craniosynostosis patients with SVs containing the Histone deacetylase 9 (HDAC9) protein coding sequence are associated with disruption of *TWIST1* regulatory elements that reside within *HDAC9* sequence. Using epigenetic marks and *in vivo* enhancer assays, we characterized six craniofacial *TWIST1* enhancers located in the *TWIST1-HDAC9* locus. Based on SVs within the *HDAC9-TWIST1* locus, we defined the 3' *HDAC9* sequence (~500Kb) as a critical Twist1 regulatory region. By deleting Twist1 enhancers within the Hdac9 protein coding sequence in mice (eTw5-7$^{Del/Del}$), we showed that *Twist1* expression was decreased, resulting in smaller sized and asymmetric skull and polydactyly. Furthermore, deletion of a Ctcf site (Ctcf$^{Del/Del}$) within the *Hdac9* protein coding sequence, disrupted *Twist1* enhancer-promoter interactions and altered Twist1 expression which led to deformed skull and hindlimb polydactyly, resembling Twist1+/- mouse phenotype. Deletions of Twist1 regulatory elements altered the distinct anteriorposterior expression patterns of Shh pathway genes, including *Hand2* and *Alx4*. Using UMI-4C, we demonstrated that both enhancers and Ctcf site regions interact with *Twist1* promoter region. These interactions are depended on the presence of both regulatory regions, indicating a specific chromatin conformation of *Hdac9* in regulating *Twist1* expression. Finally, a large inversion of the entire *Hdac9* sequence (Hdac9$^{INV/+}$) that does not disrupt *Hdac9* expression but rather repositions *Twist1* regulatory elements showed a decrease in Twist1 expression that led to subtle craniofacial phenotype and hindlimb polydactyly. Thus, our study elucidated essential components of *TWIST1* transcriptional machinery that reside within the *HDAC9* sequence, suggesting that SVs, encompassing protein coding sequence, such as HDAC9, could lead to a phenotype that is not attributed to its protein function but rather to a disruption of the transcriptional regulation of a nearby gene, such as TWIST1.

PrgmNr 1265 - De-Novo Mutations identified in Nonsyndromic Cleft lip/Palate Families from Africa

View session detail

Author Block: **W. Awotoye**[1,2], P. Mossey[3], J. B. Hetmanski[4], L. Gowans[5], M. A. Eshete[6], W. L. Adeyemo[7], A. Alade[2,8], E. Zeng[9], C. Comnick[9], C. Pendleton[9], O. Adamson[10], T. Naicker[11], C. Adeleke[2], T. Busch[2], M. Li[2], A. Petrin[12], A. Oladayo[13], S. Kayali[2], J. Olotu[14], M. Hassan[2], J. Pape[2], P. Donkor[5], T. H. Beaty[15], M. Taub[16], M. L. Marazita[17], A. A. Adeyemo[18], J. C. Murray[19], A. Butali[13]; [1]Iowa Inst. for Oral Hlth.Res., Univ. of Iowa, Iowa, IA, [2]Dept. of Oral Pathology, Radiology and Med., Coll. of Dentistry, Univ. of Iowa, Iowa, IA, [3]Univ. of Dundee, Dundee, United Kingdom, [4]JH Bloomberg Sch PH, Baltimore, MD, [5]Komfo Anokye Teaching Hosp. and Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [6]Addis Ababa Univ., Addis Ababa, Ethiopia, [7]Dept. of Oral and Maxillofacial Surgery, Coll. of Med. of the Univ. of Lagos/ Lagos Univ. Teaching Hosp. (CMUL/LUTH), Lagos, Nigeria, [8]Dept. of Epidemiology, Coll. of Publ. Hlth., Univ. of Iowa, Iowa, IA, [9]Div. of Biostatistics and Computational Biology, Coll. of Dentistry, Univ. of Iowa, Iowa, IA, [10]Dept. of Oral and Maxillofacial Surgery, Coll. of Med. of the Univ. of Lagos (CMUL), Lagos, Nigeria, Lagos, Nigeria, [11]Univ. of KwaZulu-Natal, Durban, South Africa, [12]Iowa Inst. for Oral Hlth.Res., Univ. of Iowa, Iowa City, IA, [13]Dept. of Oral Pathology, Radiology and Med., Coll. of Dentistry, Univ. of Iowa, Iowa City, IA, [14]Dept. of Anatomy, Univ. of Port Harcourt, Port Harcourt, Nigeria, [15]Johns Hopkins Univ, Sch PubHlth, Baltimore, MD, [16]Johns Hopkins, Baltimore, MD, [17]Univ Pittsburgh, Pittsburgh, PA, [18]NIH, Bethesda, MD, [19]Dept. of Pediatrics, Univ. of Iowa, Iowa City, IA

Disclosure Block: **W. Awotoye:** None.

Background: Despite successes in the investigation of *de novo* mutations (DNMs) in the etiology of some birth defects (autism, congenital heart defects), only a limited number have been reported for nonsyndromic cleft lip with or without cleft palate (NSCL/P), the most common craniofacial birth defect. To identify high impact DNMs controlling risk of NSCL/P, we conducted whole genome sequencing (WGS) analyses of case-parent trios from an understudied population.

Method: A total of 150 nsCL/P African case-parent trios were sequenced for this study. Each trio comprises an affected child (with nsCL/P) and unaffected parents and were recruited from Ghana and Nigeria. Saliva samples were collected from these individuals and their genomes were sequenced from extracted DNA. After quality control, we screened the genomes of the remaining 130 trios for high impact DNMs possibly contributing to risk of nsCL/P. We used bioinformatic prediction tools to identify those mutations predicted to damage and affect the protein structures and functions.

Results: We identified 110 potential pathogenic DNMs. These include novel loss of function (LOF) variants in *TTN, MINK1* and *ARHGAP10* genes; and missense variants in *DHRS3, TULP4, SHH, TP63, FKBP10, ACAN, RECQL4* and *KMT2D*. These variants are predicted to be damaging and are among the most deleterious (top 1%) mutations in the human genome. Experimental evidence in published works showed *TTN, SHH, TP63, FKBP10, ACAN, RECQL4* and *KMT2D* genes are involved in facial development and are involved in the etiology of syndromic CL/P. While *DHRS3, SHH* and *TP63* contribute to the risk of nsCL/P. Interestingly, our *SHH de novo* variant p.Ser362Leu has been reported to cause holoprosencephaly 3 (HPE3), a syndromic form of CL/P. Damaging mutations in the *DHRS3* gene affects retinoic acid signaling during embryogenesis which causes cleft palate. Association studies have identified *TULP4* as a potential cleft candidate gene, while *ARHGAP10* interacts with *CTNNB1* to control WNT signaling. *MINK1* plays a role in cell-cell adhesion and migration, and causes abnormal tooth morphogenesis in mice. Our gene-set enrichment analysis identified additional genes that are important in palatal development. These include *DLX6* and *EPHB2* and they both harbored novel damaging DNMs.

Conclusion: Our WGS adds to the available data on Africa population (a historically underrepresented

group in genetics study) and has identified novel pathogenic *de novo* variants that may contribute to the developmental pathogenesis of NSCL/P. These findings demonstrate the power of WGS analysis of trios for discovering potential pathogenic variants.

PrgmNr 1266 - Identification of novel molecular pathways in syndromic orofacial clefting

View session detail

Author Block: **K. Wilson**[1], D. Newbury[2], U. Kini[1,3]; [1]Oxford Ctr. for Genomic Med., Oxford Univ. Hosp. NHS Fndn. Trust, Oxford, United Kingdom, [2]Faculty of Hlth.and Life Sci., Oxford Brookes Univ., Oxford, United Kingdom, [3]Spires Cleft Ctr., John Radcliffe Hosp., Oxford, United Kingdom
Disclosure Block: **K. Wilson:** None.
Background: Syndromic orofacial clefting (OC) accounts for 30% of cleft lip and/or palate. An updated review of molecular pathways associated with syndromic OC is unavailable. The Deciphering Developmental Disorders (DDD) study provides a source of quality data to assemble this information. The Genomics England PanelApp is a publicly available database of curated virtual gene panels and is a valuable reference tool for genes associated with syndromic OC.
Aim: To investigate molecular pathways associated with syndromic OC by reviewing the results of exome sequencing (ES) and exon-arrayCGH in a large cohort of patients with syndromic OC.
Methods: Patients with the HPO terms â□□cleftâ□□ and â□□bifid uvulaâ□□ were identified through a Complementary Analysis Project within the DDD study. Possible diagnostic variants were identified by automated variant filtering and manual review. Single nucleotide variants (SNVs) within known disease-causing genes and copy number variants (CNVs) were classified according to the ACMG guidelines, the ACGS Best Practice Guidelines and consensus opinion. Functional analyses of identified genes were performed within STRING, Cytoscape and MCODE. Associated phenotypes were explored using the International Mouse Phenotyping Consortium. Gene expression analyses were performed within GENE2FUNC.
Results: 603/13612 (4.4%) patients were identified of whom 453/603 (75.1%) had trio ES. Pathogenic (P) or likely pathogenic (LP) variants were identified for 220/603 (36.5%) patients in 124 known disease-causing genes with *SATB2* the most common (16/220, 7.3%). 23/220 (10.5%) patients had a P or LP CNV of partial or full contribution to their phenotype. 35/124 genes fulfilled criteria to be added to the PanelApp â□□Cleftingâ□□ panel, increasing the size of the current panel by 23.8%. Gene ontology and pathway analyses identified novel molecular networks for syndromic OC which were distinct from those in non-syndromic OC. Gene expression analyses and investigation of knockout phenotypes also showed a distinction between syndromic and non-syndromic OC. Pathway and expression analyses showed an enrichment of genes associated with intellectual disability (FDR=2.8x10-33), RNA metabolism (FDRConclusion: This study demonstrates the utility of ES and CNV analysis for patients with syndromic OC and increases the diagnostic rate for this patient cohort. It also highlights novel molecular pathways specific to syndromic OC and enhances our understanding of lip and palate development.

PrgmNr 1267 - RERE deficiencycontributes to the development of orofacial clefts in humans and mice

View session detail

Author Block: **B-J. Kim**[1], H. Zaveri[1], V. K. Jordan[2], P. N. Kundert[1], T. M. Scott[3], J. Carmichael[4], D. A. Scott[5]; [1]Molecular & Human Genetics, Baylor Coll. Med., Houston, TX, [2]Baylor Coll. Med., Houston, TX, [3]Brigham Young Univ., Provo, UT, [4]Northampton Gen. Hosp., Cliftonvile, United Kingdom, [5]Baylor Coll. Med, Houston, TX
Disclosure Block: **B. Kim:** None.
Deletions of chromosome 1p36 are the most common telomeric deletions in humans and are associated with an increased risk of orofacial clefting. Deletion/phenotype mapping, combined with data from human and mouse studies, suggests the existence of multiple 1p36 genes associated with orofacial clefting including *SKI*, *PRDM16*, *PAX7*, and *GRHL3*. The arginine-glutamic acid dipeptide (RE) repeats gene (*RERE*) is located in the proximal critical region for 1p36 deletion syndrome and encodes a nuclear receptor co-regulator. Pathogenic *RERE* variants have been shown to cause neurodevelopmental disorder with or without anomalies of the brain, eye, or heart (NEDBEH), but have not been shown to cause orofacial clefting. Here we report the first individual with NEDBEH to have a cleft palate. We confirm that RERE is broadly expressed in the palate during mouse embryonic development, and we demonstrate that the majority of RERE-deficient mouse embryos on C57BL/6 background have cleft palate. We go on to show that ablation of *Rere* in cranial neural crest cells, mediated by a Wnt1-Cre, leads to delayed elevation of the palatal shelves and cleft palate, and that proliferation of mesenchymal cells in the palatal shelves is significantly reduced in $Rere^{flox/flox}$;Wnt1-Cre embryos. We conclude that loss of RERE function contributes to the development of cleft palate in individuals with proximal 1p36 deletions and NEDBEH, and that RERE expression in cranial neural crest cells and their derivatives is required for normal palatal development.

PrgmNr 1268 - Single cell transcriptomics-directed investigation of *de novo* mutations and rare inherited genetic variants in cleft lip and palate

View session detail

Author Block: **K. Robinson**[1], M. Losa[2], K. K. Diaz Perez[1], S. W. Curtis[3], L. M. Moreno Uribe[4], T. H. Beaty[5], J. C. Murray[6], E. Feingold[7], M. L. Marazita[8], L. Selleri[9], E. J. Leslie[1]; [1]Emory Univ., Atlanta, GA, [2]Univ. of California San Francisco, San Francisco, CA, [3]Emory Univ., Tucker, GA, [4]Univ Iowa, Iowa City, IA, [5]Johns Hopkins Univ, Sch PubHlth, Baltimore, MD, [6]U of Iowa, Iowa City, IA, [7]Univ of Pittsburgh, Pittsburgh, PA, [8]Univ Pittsburgh, Pittsburgh, PA, [9]Univ. of California San Francisco, San Franciso, CA
Disclosure Block: **K. Robinson:** None.
Cleft lip and palate (CL/P) is one of the most common congenital anomalies. The etiology of CL/P is complex with both environmental and genetic risk factors. While previous studies have identified several CL/P-associated genes or regions, only a fraction of all cases can be clearly attributed to specific genes. Additional genetic causes may be due to rare inherited variants (RIVs) or *de novo* mutations (DNMs) in simplex CL/P cases. To investigate this further, we performed single cell RNA sequencing on epithelial cells of the lambdoidal junction (Î») from gestational day (E)10.5 wildtype mouse embryos at the point of upper lip fusion. We identified six cell clusters, and using the top 150 differentially expressed genes from each, we carried out targeted evaluation of both DNMs and RIVs in whole genome sequences from 756 CL/P case-parent trios of Asian, Latino, and European ancestry. For each cluster we analyzed enrichment of nonsynonymous, loss-of-function (LOF), and protein altering (nonsynonymous + LOF) DNMs. Of these, the olfactory epithelium cluster was enriched for protein altering (p=0.005) and the periderm cluster was enriched for nonsynonymous variants (p=0.005). We then evaluated exonic RIVs as defined by a minor allele frequency of IRF6, *TFAP2A*, and *GRHL3*, all of which contained both DNMs and RIVs, suggesting other genes identified via this method may also be significant in risk for CL/P. Although all clusters were evaluated, the Î»-fusion effector cluster was of specific interest given its critical role in prominence fusion during normal craniofacial development. This cluster harbored variants of interest in 31 genes, a higher percentage than other clusters (23% vs. 9-19%). Further, gene ontology revealed a group of 14 genes enriched for terms related to transcription and, interestingly, both negative regulation of epithelial cell proliferation (FDR 0.015) and positive regulation of mesenchymal cell proliferation (FDR 0.049). Among this group was transcription factor *ZFHX3*, which contained the most variants including loss-of-function DNMs and RIVs; thus, it remains of high interest for novel CL/P risk association. This investigation illustrates the benefit of integrating genomic technologies to prioritize and identify novel genetic associations with risk to CL/P. Continued evaluation focused on gene interactions and pathways represented by these variants may further elucidate CL/P etiology.

PrgmNr 1269 - A novel 3 MB deletion on 6p24 removes distant neural-crest enhancers controlling *TFAP2A* resulting in mild Branchiooculofacial syndrome in a multiplex family with orofacial clefting

View session detail

Author Block: **T. Yankee**[1], C. Willett[2], K. Child[1], A. Petrin[3], A. Wilderman[4], T. Cox[5], T. Roscioli[6], L. Moreno-Uribe[3], E. J. Leslie[7], J. Cotney[8,9]; [1]Dept. of Genetics and Genome Sci., UConn Hlth.Graduate Program in Genetics and Dev.al Biology, Farmington, CT, [2]Emory Univ. Sch. of Med. Dept. of Genetics, Atlanta, GA, [3]Univ. of Iowa, Iowa City, IA, [4]UConn Hlth.Graduate Program in Genetics and Dev.al Biology, Farmington, CT, [5]Univ. of Missouri-Kansas City, Kansas City, MO, [6]Univ. of New South Wales, Sydney, Australia, [7]Emory Univ., Atlanta, GA, [8]Univ. of Connecticut Sch. of Med., Farmington, CT, [9]Univ. of Connecticut Inst. for Systems Genomics, Farmington, CT

Disclosure Block: **T. Yankee:** None.

In a previous effort to characterize large copy number variants as genetic risk factors for orofacial clefting (OFC) following a genome-wide association study, we identified a novel 3 Mb deletion on 6p24 in three affected members of the same family from Colombia. The reported pedigree had a dominant inheritance pattern and included 5 individuals with OFC. All affected family members carried the deletion, which was inherited from the probandâ□□s unaffected grandmother. Supporting the pathogenicity of this deletion, exome sequencing of this family found no segregating single nucleotide variants in OFC candidate genes. The 3Mb deleted region included 12 genes, none of which are strong OFC candidates and a 840 kb gene desert. We observed that the 3â□□ breakpoint of this deletion occurred within a large non-coding region directly adjacent to the important developmental transcription factor *TFAP2A*. *TFAP2A* loss of function is one of the primary genes associated with Branchiooculofacial syndrome (BFOS) that includes OFC in conjunction with branchial and ocular abnormalities. Affected members of this family display OFC, broad nasal root, and slight hypotelorism characteristic of BFOS. Cell type-specific enhancers in the genomic region directly flanking *TFAP2A* have been previously implicated in a subset of BOFS patients, however these enhancers remain intact in this family.We hypothesized that this deletion contains additional regulatory elements of *TFAP2A*, resulting in attenuated expression of this gene and a milder form of BOFS. To address this hypothesis and predict enhancer-gene interactions at this region, we used an activity by contact model (ABC-Enhancer) to integrate ChIP-seq based chromatin state annotations, chromosome conformation (Hi-C), and gene expression (RNA-Seq) from primary human craniofacial tissues and a culture model of cranial neural crest cells (CNCCs). We identified closely clustered, strong enhancer states that were predicted to have interactions with *TFAP2A* over a distance of greater than two megabases. To validate these predictions *in vivo*, we used CRISPR-Cas9 in human embryonic stem-cells to create homozyguous deletions of a 25kb region encompassing the strong enhancer segments. We differentiated these lines to CNCCs and compared gene expression and proliferative capacity to wildtype. We found this region is essential for cell viability during CNCC differentiation and differential gene expression analysis revealed substantial effects on *TFAP2A* expression. In summary, we identified a region within the inherited deletion that regulates *TFAP2A* gene expression and could contribute to the mild BOFS phenotype in this family.

PrgmNr 1272 - Increasingâ¯Detection ofâ¯Familial Hypercholesterolemiaâ¯in an Academic Medical Center:â¯Theâ¯Familial Hypercholesterolemia Identification Registryâ¯(FHIRE).â¯

View session detail

Author Block: **A. Miller**, S. Nigbur, I. J. Kullo; Mayo Clinic, Rochester, MN
Disclosure Block: **A. Miller:** None.
Background:â¯Familial Hypercholesterolemia (FH) is a common genetic disorder with a prevalence of 1:2201. FH is a Tier 1 genetic disorder, and early detection and treatment of FH has potential to positively impact public health2.â¯ There is no universal screening mechanism for FH in the United States, and it is estimated that 90% of people with FH are undiagnosed3.â¯ Methods:â¯â¯The Familial Hypercholesterolemia Identification Registry (FHIRE)â¯combines an automatedâ¯algorithm withâ¯manual review to identify patients with FHâ¯at Mayo Clinic Rochester. Daily lists of patients who have anâ¯LDL-cholesterol >190 mg/dL and an upcoming appointment areâ¯generatedâ¯in an automated fashionâ¯and sentâ¯to the study team. Theâ¯electronic health recordsâ¯of theseâ¯patients areâ¯manually reviewed to calculate a Dutch Lipid Clinic Network (DLCN) or MEDPED score. Patients with a DLCN scoreâ¯ofâ¯â¯â¥6 or who met MEDPED criteria for definite FH areâ¯offered enrollmentâ¯in theâ¯registry. Patients who consent undergo CLIA certified genetic sequencing ofâ¯LDLR,â¯APOB, andâ¯PCSK9.â¯ Results:â¯â¯Between 1/1/2019 and 6/8/2021,â¯95â¯participantsâ¯consentedâ¯andâ¯84â¯completedâ¯geneticâ¯testing. Of these,â¯38â¯(45.2%)â¯hadâ¯aâ¯pathogenic/likely pathogenicâ¯(P/LP)â¯variant consistent with a diagnosis ofâ¯monogenicâ¯FH;â¯9â¯carried the APOB founder variant c.10580G>A (p.Arg3527Gln)â¯andâ¯1 carriedâ¯theâ¯c.10579C>Tâ¯(p.Arg3527Trp)â¯P/LPâ¯APOB variant.â¯The remainingâ¯28â¯had P/LP variants in LDLR:â¯18â¯missenseâ¯(includingâ¯1â¯in-frame insertion/deletion),â¯1 duplication,â¯5â¯copy number variants, andâ¯5â¯splice site alterations.â¯8â¯participants (9.5%) hadâ¯aâ¯variant of uncertain significanceâ¯(VUS);â¯5â¯had both a P/LP and VUSâ¯(6%),â¯andâ¯38â¯(45.2%)â¯had negative findings.â¯The detection rateâ¯of monogenic FH inâ¯FHIREâ¯was 56%.â¯Participants with a P/LP and/or VUS finding were invited to discuss their results with a genetic counselor.â¯â¯ Conclusion:â¯We designedâ¯a novel, two-step protocol for identifying individuals with monogenic FHâ¯in an academic medical center. The protocol has a high detection rate and could be deployed across health care systems to increase the detection of FH.â¯

PrgmNr 1273 - Finding the Sweet Spot: Exploring patientsâ perspectives on the adoption of AI and Chatbots in genetic service delivery

View session detail

Author Block: **Y. Bombard**[1,2], S. Luca[3], M. Clausen[2], A. Shaw[2], W. Lee[1,3], S. Krishnapillai[1,2], E. Adi-Wauran[1,2], H. Faghfoury[4], G. Costain[5], R. Jobling[5], M. Aronson[4,6], E. Liston[5,7], J. Silver[4,7], C. Shuman[7], L. Chad[5], R. Z. Hayeems[1,3]; [1]Inst. of Hlth.Policy, Management, and Evaluation, Univ. of Toronto, Toronto, ON, Canada, [2]Genomics Hlth.Services Res. Program, Li Ka Shing Knowledge Inst., St. Michaelâs Hosp., Unity Hlth.Toronto, Toronto, ON, Canada, [3]Program in Child Hlth.Evaluative Sci., The Hosp. for Sick Children, Toronto, ON, Canada, [4]Fred A. Litwin Family Ctr. in Genetic Med., Univ. Hlth.Network and Sinai Hlth.System, Toronto, ON, Canada, [5]Div. of Clinical and Metabolic Genetics, The Hosp. for Sick Children, Toronto, ON, Canada, [6]Zane Cohen Ctr. for Digestive Diseases, Sinai Hlth.System, Toronto, ON, Canada, [7]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **Y. Bombard:** None.
Background: Innovative strategies are needed to increase the capacity and efficiency of genetic service delivery in light of workforce shortages. eHealth tools such as digital platforms, artificial intelligence (AI) and conversational agents, such as chatbots, are potential solutions. Digitizing components of care can increase the time available for in-person patient care, improve access and potentially decrease administrative burden on providers. However, research exploring patientsâ preferences and attitudes towards AI and chatbots in genetic service delivery is only emerging.
Aim: To understand patient perspectives on digitizing the genetic counseling and testing experience through a patient-facing genetics platform that integrates AI and a chatbot.
Methods: A qualitative study was conducted using semi-structured interviews to explore patient perspectives on a proposed genetics platform integrating AI and a chatbot. Thematic analysis employing interpretive description was used to analyze interview transcripts.
Results: We interviewed 30 participants (n= 20 female; age range= 30-79, n= 17 patients, n=13 parents of patients) who previously received genetic testing for themselves or their child. Participants perceived AI as an opportunity to provide them with âpersonalizedâ information. However, they expressed reservations about the use of chatbots. Specifically, they anticipated that chatbots would be âinefficientâ or lack the capabilities needed to manage complex interactions but said they may be helpful in providing some types of basic information. Participants expressed that the utility of a chatbot would be low if it were designed to assist with tasks that are either too simple (e.g., defining medical terminology) or too complex (e.g., responses incorporating personal health information). The highest level of acceptability was in the âsweet spotâ, where participants saw value in a chatbot that could assist with moderately complex tasks (e.g., providing a pros/cons list for a specific type of genetic testing). Irrespective of the chatbotâs complexity, participants needed a âsafety netâ in the form of access to a care provider throughout their use of the platform, in case the chatbot was unable to address their needs. Knowing that a care provider would be available to them as needed increased participantsâ comfort with using chatbots or AI as part of the platform.
Conclusion: Results provide timely insights into patientsâ comforts with and limits on integrating AI and chatbots into genomics platforms, which can optimize implementation of these tools in practice.

PrgmNr 1274 - A Step Towards Real-World Population Screening for Hereditary Cancer and Hypercholesterolemia

View session detail

Author Block: **B. H. Shirts**[1], N. Rao[2], J. Huey[1], S. Coe[1], D. Tsinajinnie[1], J. Kaganovsky[1], G. P. Jarvik[3], S. M. Fullerton[2], A. T. Chen[1]; [1]Univ of Washington, Seattle, WA, [2]Univ. of Washington, Seattle, WA, [3]Univ Washington Med Ctr., Seattle, WA

Disclosure Block: **B.H. Shirts:** None.

Timely identification of individuals carrying risk variants for Hereditary Breast and Ovarian Cancer Syndrome, Lynch Syndrome, and Hereditary Hypercholesterolemia is vital for treating, delaying or preventing disease onset. Population genetic screening is a proposed strategy to detect individuals who would benefit from medical intervention but who are currently missed under existing screening guidelines.

We sought to assess the feasibility of population screening by inviting adults who have received care at the University of Washington Medical Center to participate in a targeted hereditary disease screening. Unlike other population screening studies, our study does not focus on individuals who have previously enrolled in genetic research. We attempted to design a study that would be as similar as possible in approach to a member of the general public requesting a COVID-19 testing appointment online. The study started in June 2020. Adults were identified through the electronic health record regardless of family or personal history of disease, oversampling for racial and ethnic minorities as well as LGBTQ+ individuals (diverse participants). After they receive an email invitation to participate, invited individuals can learn more about the study by clicking a link in their email invitation to the study FAQ, which includes details about the studyâ□□s targeted testing, minimal health education about genetics, potential risks of genetic screening, and additional information about study participation.

Participants could request a DNA saliva kit be mailed to them for sample collection. Samples returned were screened for pathogenic or likely pathogenic variants in 25 genes associated with increased risk of cancer, including *BRCA1/2* and genes associated with Lynch, or familial hyperlipidemia using a custom capture panel and sequencing. Participants who tested positive for a pathogenic or likely pathogenic variant receive counseling about results from study genetic counselors and are advised to receive follow-up care with genetics providers. All other participants are able to access their results through a secure online portal.

Thus far, more than 38,000 people have been invited for genetic screening and >1,700 (77% racial, ethnic minorities, and/or LGBT+) have chosen to participate. We will reach our goal of 2,500 participants in summer 2021. Positive results were observed in 4.6% of participants tested to date; almost half were already aware of their genetic risk. Full results from this population screening study will help determine if broad implementation of genetic screening is an effective way to identify diverse at-risk individuals.

PrgmNr 1275 - Using Supervised Machine Learning to Accelerate Variant Curation and Genomic Diagnosis in Suspected Mendelian Disorders

View session detail

Author Block: **P. Dai**[1,2], R. Fieldhouse[3], A. Honda[4], J. Copty[4], A. Statham[4], M. Hobbs[4], T. Ohnesorg[4], G. Hollway[2,4], B. A. Lundie[4], E. Lee[4], K. Enthoven[1], S. Tangye[1], P. Gray[5], M. Wong[5], W. Kaplan[2,4], T. Phan[1,2], L. Burnett[2,4]; [1]Garvan Inst. of Med. Res., Sydney, Australia, [2]St Vincent's Clinical Sch., Faculty of Med., Univ. of New South Wales, Sydney, Australia, [3]Univ. of Sydney, Sydney, Australia, [4]Kinghorn Ctr. for Clinical Genomics, Sydney, Australia, [5]Sydney Children's Hosp. Network, Sydney, Australia
Disclosure Block: **P. Dai:** None.
Background: Massive parallel sequencing (MPS) has revolutionized the diagnosis of rare Mendelian disorders over the last decade. However, despite technological improvements in sequencing and bioinformatics, there still exist significant bottlenecks in variant curation due to the need for manual review of large number of potential variants against in silico annotations, the literature and clinical phenotype. **Aim:** We developed a parsimonious algorithm for rapid variant prioritization using supervised machine learning (ML) combined with an orthogonal phenotype-based variant prioritization tool. **Method:** We generated a data set of 158,626 genomic variants drawn from the genomes of 291 near-consecutive patients who underwent diagnostic whole genome sequencing (WGS) at the Kinghorn Centre for Clinical Genomics (KCCG) between 2015 and 2018. We subsequently extracted 121 American College of Medical Genetics and Genomics (ACMG) Class 4/5 (likely pathogenic/pathogenic) variants and an equal number of randomly selected variants that had been deemed as non-diagnostic by ACMG criteria. Using the â□□Wekaâ□□ ML environment (University of Waikato, New Zealand), we subsequently performed attribute reduction and developed a variant prioritization ML parsimonious algorithm based on in silico genomic annotations by training the J48 algorithm on this dataset with diagnostic performance assessed by 10-fold cross validation. We also separately evaluated a variant prioritization orthogonal algorithm based purely on patient phenotypes (AMELIE, Stanford University, CA USA) without inclusion of any genomic attributes. **Results:** The ML algorithm successfully captured 119/121 (98%) of ACMG Class 4/5 pathogenic variants with 15/121 (12%) of the non-pathogenic variants misclassified as pathogenic. Overall the ML algorithm reduced the number of variants requiring manual curation by approximately 2/3. We further applied the short-listed set of variants from the ML algorithm to AMELIE in 10 randomly selected cases and found the candidate pathogenic variant to be occurring within the top 10 short-listed variants in 100% of cases. **Conclusion**: A simple supervised machine learning algorithm trained on manually curated genomic variants in combination with an orthogonal phenotype-based variant prioritization tool can markedly reduce the workload of variant curation without significant sacrifices in sensitivity. This will allow both rapid semi-automated initial genome analysis and subsequent re-analysis of suspected Mendelian disorders. Further work will involve testing our pipeline on larger independent and prospective test data sets.

PrgmNr 1276 - Design and implementation of genome interpretation for coronary disease - including monogenic and polygenic risk assessment - in a Preventive Genomics Clinic

View session detail

Author Block: **D. Brockman**[1], D. Maamari[1], R. Pelletier[1], E. Folkerts[1], C. Leonard[1], L. Petronio[2], K. H. Huang[2], P. Natarajan[1], K. Aragam[1], H. L. Rehm[1], A. V. Khera[1]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA
Disclosure Block: **D. Brockman:** None.
Introduction: State-of-the-art genome interpretation for common diseases is now recognized as requiring assessment of both monogenic variants - single base pair changes that disrupt a specific pathway - and polygenic background, the cumulative impact of many common variants. Taking coronary artery disease (CAD) as an example, we and others have noted striking ability to stratify the population - from birth - into genetic subgroups with very different trajectories of disease onset. Alongside considerable (and warranted) enthusiasm for this new approach are several areas of uncertainty around optimal approach for implementation and disclosure.
Methods: First, we designed a web-based user educational tool focused on the scientific basis of polygenic scoring and a PDF report designed for inclusion in the medical record. Second, we partnered with Color Health to develop a clinical-grade test for both monogenic variants and polygenic score calculation. Third, we designed a prospective study to assess monogenic and polygenic drivers for CAD and evaluate impact of risk disclosure for patients seen in a new Preventive Genomics Clinic at Massachusetts General Hospital.
Results: A new educational tool created using â□□scrolly-tellingâ□□ principles (polygenicscores.org/explained/) was developed with a team of data visualization, clinical, and genetic experts - now viewed 7600 times across 19 countries. A static PDF report was drafted and refined based on formal user experience testing. We launched a clinical test - enabled by saliva collection kits sent directly to patientâ□□s homes - for CAD assessment in our clinic. Of the first 34 patients, median age was 50.5 years (range 27 to 77), 71% were male, and 74% were white. Primary rationale for obtaining the test included a strong family history (71%), a personal history of heart disease and interest in understanding any genetic contribution (32%), or concern in the absence of family or personal history (15%). The disclosed results led to a change in clinical management for 56% of patients studied to date - including informing decision to start new preventive medications, additional imaging scans or biomarkers, or stated intent to enhance adherence to a healthy lifestyle. 92% of participants reported learning valuable information related to their health and 83% reported the reporting tools were either â□□veryâ□□ or â□□extremelyâ□□ helpful.
Conclusions: We describe a generalizable approach to develop and test tools for sophisticated genetic prediction models understandable to both patient and clinical stakeholders. These results are likely to inform ongoing efforts related to genetic risk disclosure within clinical practice.

PrgmNr 1277 - Innovative approaches to optimizing the electronic health record for genomic medicine

View session detail

Author Block: **K. Lau-Min**[1], S. Byers Asher[2], D. Mckenna[1], T. Bardakjian[3], J. Chen[4], S. M. Domchek[1], J. Landgraf[4], A. Schoenbaum[4], G. Hoffecker[2], S. Tuteja[2], L. Steinfeld[5], C. Vanzandbergen[4], M. D. Ritchie[6], K. L. Nathanson[2]; [1]Div. of Hematology/Oncology, Dept. of Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [2]Div. of Translational Med. and Human Genetics, Dept. of Med., Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [3]Dept. of Neurology, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [4]Information Services Applications, Penn Med., Univ. of Pennsylvania, Philadelphia, PA, [5]Carey Law Sch., Univ. of Pennsylvania, Philadelphia, PA, [6]Dept. of Genetics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA
Disclosure Block: **K. Lau-Min:** None.
The PennChart Genomics Initiative (PGI) at the University of Pennsylvania is a multidisciplinary collaborative that aims to optimize the electronic health record (EHR) for the delivery of genomic medicine. We leverage the collective expertise of clinicians, pathologists, researchers, legal staff, and information services throughout Penn Medicine and its partners to devise innovative EHR-based approaches. We previously described our initial efforts to link our EHR (PennChart) directly with clinical testing laboratories, integrate discrete genetic data, link to clinical decision support, and enhance patient access to their own genetic testing results. The PGI has since expanded to interface with three commercial genetic testing laboratories for computerized provider order entry (CPOE) and discrete result reporting, resulting in the creation of over 390 unique genetic testing orders. The CPOE interface has also been refined to include additional details such as test specimen source, as well as linkage to patient demographic information to facilitate insurance benefits investigations. To date, a total of 1,065 genetic testing orders have been placed as a part of clinical care, with a 31% increase over the last six months. Results from these tests filter discretely into the PennChart Precision Medicine Tab, a centralized location in the EHR that enables easy access to all genetic data, including over 17,500 legacy scans containing unstructured genetic data. The number of Genomic Indicators - tags that indicate potential disease risk or drug sensitivity based on a patientâ□□s discrete genetic testing results - has also expanded to include 21 disease-associated and 55 pharmacogenetic variants, enabling easier interpretability for non-genetics clinicians and patients while also facilitating the development of downstream clinical decision support. Finally, to mitigate the confusion or distress that may arise if patients receive their genetic testing results in isolation, a Genomics Release Portal has been built to retain genetic testing results from release to the EHR until after the ordering genetics provider has had the opportunity to review, disclose, and counsel patients on their clinical interpretation. All in all, the PGIâ□□s efforts have significantly streamlined what was previously a manual process of placing genetic testing orders, handling results, and interfacing with the EHR, resulting in substantial improvements in provider workflows and satisfaction. Future work will continue to broaden the programâ□□s scope of genetic variants and clinical decision support systems to further optimize the delivery of genomic medicine for our patients.

PrgmNr 1280 - Understanding clonal distributions during human cortical development deciphered through brain somatic mosaicism

View session detail

Author Block: **X. Yang**[1,2], M. W. Breuss[3,1,2], J. C. M. Schlachetzki[1], D. Antaki[1,2], J. G. Gleeson[1,2]; [1]Univ. of California, San Diego, La Jolla, CA, [2]Rady Children's Inst. for Genomic Med., San Diego, CA, [3]Univ. of Colorado, Sch. of Med., Aurora, CO
Disclosure Block: **X. Yang:** None.
The structure of the human neocortex underlies species-specific traits and is a reflection of intricate developmental programs. Yet, lineage tracing and clonal analysis in the human brain are challenging compared with other species. Somatic mosaic mutations across the genome have the potential to address this problem when employing them as neutral recorders for lineage history and early embryonic development of the human cortex. Thus, a thorough study of the distribution of mosaic clones in human cortical areas may help us to understand several unanswered questions about the normal and malformatted brain: 1] How are clones in the human brain distributed during early embryonic development? 2] What is the total committed progenitor population size for the early human cortex? 3] Why are brain disorders such as focal cortical dysplasias (FCD) in humans limited to a single hemisphere?We addressed these questions through the study of neurotypical postmortem human brains through a comprehensive assessment of brain somatic mosaicism, which deciphered neocortical cellular lineages. By combining a sampling of 50 distinct anatomic locations with 300x ultra-deep whole-genome sequencing from 4 neurotypical cadavers, we identified hundreds of *bona fide* mosaic variants, resolving major cortical cell types, and single-nuclei clade organizations across the brain and other organs.We found that clones derived after the accumulation of 90 to 200 progenitors in the cerebral cortex tended to respect the midline axis, well before clones respect anterior-posterior or ventral-dorsal axes, and that cell fate correlated better with location rather than early lineage. There was evidence for local clonal expansion as well as a broad range of stochastic distributions of somatic variants across the cortex. Similar to rodents, cortical cells originated from both dorsal and ventral cellular populations, whereas microglia lineages appeared as distinct from other resident brain cells.Our data provide a comprehensive analysis of brain somatic mosaicism across the cortex, demonstrate cellular origins and migration patterns within typically developed human brains, and reveal mechanisms how mosaic variants lead to focal brain disease. Future analysis can help to deconvolve cellular lineages.

PrgmNr 1281 - *MYCN* gain-of-function variants induce excess proliferation of neurons and cause a novel megalencephaly syndrome

View session detail

Author Block: **Y. Nishio**[1,2], K. Kato[1,2,3], T. M. Frederic[4], S. Otsuji[1], C. QuÃ©lin[5], H. Oishi[6], Y. Takahashi[2], S. Saitoh[1]; [1]Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, [2]Dept. of Pediatrics, Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, [3]Dept. of Genetics, Res. Inst. of Environmental Med., Nagoya, Japan, [4]UnitÃ© Fonctionnelle 6254 dâInnovation en Diagnostique GÃ©nomique des Maladies Rares, PÃ´le de Biologie, CHU Dijon Bourgogne, Dijon, France, [5]Service de GÃ©nÃ©tique Clinique, CLAD Ouest, CHU Rennes, HÃ´pital Sud, Rennes, France, [6]Dept. of Comparative and Experimental Med., Nagoya City Univ. Graduate Sch. of Med. Sci. and Med. Sch., Nagoya, Japan

Disclosure Block: **Y. Nishio:** None.

Background: *MYCN,* a member of the *MYC* proto-oncogene family, encodes a transcription factor that regulates genes promoting cell growth and proliferation. Loss-of-function variants in *MYCN* are the causes of Feingold syndrome characterized by microcephaly, brachydactyly, and syndactyly. In contrast, we previously reported a patient with a novel megalencephaly syndrome with ventriculomegaly, neuroblastoma, and polydactyly, who carried a gain-of-function variant in *MYCN* (p.Thr58Met). Although it was likely that gain-of-function variants cause a mirror phenotype of Feingold syndrome, additional cases were needed to confirm it and elucidate the underlying molecular mechanism. **Methods:** Exome sequencing and cellular biological analyses were performed to identify and evaluate a pathogenic variant. We also generated knock-in mice ($Mycn^{WT/T58M}$) and knock-out mice ($Mycn^{WT/del}$) to delineate the pathophysiological significance of the gain-of-function or loss-of-function variants for fetal development. **Results:** We identified a *de novo MYCN* missense variant (p.Pro60Leu) in a fetus delivered under artificial abortion at the 32nd week of gestation. The patient exhibited megalencephaly, ventriculomegaly and postaxial polydactyly, that were similar phenotypes to the previous one. Immunoblot analyses of MYCN-P60L indicated a drastically decreased phosphorylation level at T58, whose phosphorylation lead to MYCN protein degradation. In fact, MYCN-P60L showed significantly increased stability and excess accumulation, indicating the gain-of-function properties as MYCN-T58M. The excess accumulation of Mycn protein was also confirmed in the E14.5 brain cortex of $Mycn^{WT/T58M}$ mice. In addition, as with the phenotypes of the patient, a postaxial polydactyly and macrocephaly with increased number of neuronal cells were observed in $Mycn^{WT/T58M}$ mice compared to littermate controls in P56, whereas $Mycn^{WT/del}$ showed microcephaly. **Conclusions:** These results confirm that *MYCN* gain-of-function variants cause a novel megalencephaly syndrome. Mirror phenotypes of Mycn gain-of-function and loss-of-function mice model suggest that MYCN plays a crucial proliferative role during neuronal development.

PrgmNr 1282 - Gain and loss of function variants in *EZH1* disrupt neuronal differentiation and lead to overlapping neurodevelopmental disorders

View session detail

Author Block: **N. Akizu**[1,2], C. Gracia-Diaz[1,2], Y. Zhou[1], G. Otrimski[1], C-H. Lee[3], P. Espana-Bonilla[4], N. Padilla[5], D. Li[1], S. Sheppard[1], P. Mark[6], M. H. Harr[1], H. Hakonarson[1], L. Rodan[7], A. Jackson[8], S. Mohammed[9], E. Concepcion[10], A. Hurst[11], M. L. Thompson[12], A. Chassevent[13], C. L. Smith-Hicks[14], M. Saade[4], X. De la Cruz[5], D. Reinberg[15], A. Holtz[7], S. Banka[16], F. S. Alkuraya[17], P. B. Agrawal[7], M. Martinez-Balbas[4], E. J. Bhoj[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Seoul Natl. Univ., Seoul, Korea, Republic of, [4]IBMB, Barcelona, Spain, [5]Hosp. Vall dâ□□Hebron, Barcelona, Spain, [6]Spectrum Hlth., Grand Rapids, MI, [7]Boston Children's Hosp., Boston, MA, [8]Teaching Hosp. NHS Trust, Blackpool, United Kingdom, [9]Guy's Hosp., London, United Kingdom, [10]Temple Univ., Philadelphia, PA, [11]Univ. of Alabama at Birmingham, Birmingham, AL, [12]Hudson Alpha Inst. for Biotechnology, Huntsville, AL, [13]NIH - NHGRI, Bethesda, MD, [14]Kennedy Krieger Inst./Johns Hopkins Univ SOM, Baltimore, MD, [15]HHMI/NYU Langone Sch. of Med., New York, NY, [16]Univ. of Manchester, Sale, United Kingdom, [17]Dept. of Translational Genomics, Ctr. for Genomic Med., King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia

Disclosure Block: **N. Akizu:** None.

Enhancer of Zeste Homologue 1 (*EZH1*) encodes one of the two Histone H3 Lysine 27 (H3K27) methyltransferases of the Polycomb Repressive Complex 2 (PRC2). Its paralogue, EZH2, has been traditionally considered the major responsible of the H3K27 trimethylation (H3K27me3) involved in the PRC2 mediated transcriptional silencing. However, EZH2 is mostly expressed in dividing cells, while EZH1 is widely expressed in both dividing and postmitotic cells, including in the developing and adult nervous system where it also contributes to H3K27me3. Whereas the function of EZH2 has been extensively characterized and its dysregulation associated with cancer and development disorders (i.e. Weaver syndrome), reports implicating EZH1 in human disorders are scarce and its biological relevance poorly understood. Here we report heterozygous missense and biallelic truncating variants in *EZH1* in ten children with neurodevelopmental delay. Recurrent clinical features of these children include intellectual disability, delayed developmental milestones, and dysmorphic facial features. By western blot analysis of a lymphoblastoid cell line derived from one of five patients, we show that biallelic truncating mutations lead to loss of EZH1 expression. In contrast, *in vitro* methyltransferase assays with three of the five missense variants showed modified methyltransferase kinetics that led to increased H3K27me3 when the variants were expressed in neural stem cells. To test if mutations affect neural development, we generated human pluripotent stem cells (hPSCs) that carry EZH1 loss or gain of function variants ($EZH1^{LOF}$ and $EZH1^{GOF}$) and differentiated them to neurons. EZH^{LOF} hPSCs showed delayed neuronal differentiation and shorter neurites, while $EZH1^{GOF}$ hPSCs exhibited premature differentiation and longer neurites. Finally, we employed the chick embryo neural development model to test the effect of EZH1 loss or gain of function *in vivo*. Consistent with the hPSC results, knocking down EZH1 in the chick embryo neural tube led to a reduced number of postmitotic neurons while its overexpression induced a premature differentiation of neural progenitor cells. These results reveal a previously unreported function of EZH1 in neuronal differentiation and uncover overlapping neurodevelopmental disorders associated with gain and loss of function variants in EZH1.

PrgmNr 1283 - Genomic atlas of proteome from three tissues relevant to neurological disorders prioritizes proteins implicated in brain diseases

View session detail

Author Block: **C. Cruchaga**; Washington Univ. St. Louis, St. Louis, MO
Disclosure Block: **C. Cruchaga:** None.
Understanding the tissue-specific genetic architecture of protein levels is instrumental to understand the biology of health and disease. We generated a genomic atlas of protein levels in multiple neurologically relevant tissues (380 brain, 835 cerebrospinal fluid (CSF) and 529 plasma), by profiling thousands of proteins (713 CSF, 931 plasma and 1079 brain) in a large and well-characterized cohort. We identified 274, 127 and 32 protein quantitative loci (pQTL) for CSF, plasma and brain respectively. cis-pQTL were more likely to be shared across tissues but trans-pQTL tend to be tissue-specific. Between 44% to 68.2% of the pQTL do not colocalize with expression, splicing, methylation or histone QTLs, indicating that protein levels have a different genetic architecture to those that regulate gene expression. By combining our pQTL with Mendelian Randomization approaches we identified potential novel biomarkers and drug targets for neurodegenerative diseases including Alzheimer disease and frontotemporal dementia. We identified one CSF, 13 plasma, and six brain proteins were likely to be in the causal pathways for AD risk. Among these proteins, plasma CD33 was a risk factor towards AD and is a drug target for other diseases, such as prostate cancer. As for Parkinson disease risk, 13 CSF, 12 plasma, and 23 brain proteins were likely to be the cause. Among these proteins, plasma IDUA was prioritized as a risk locus encoded it for PD and as a drug target for chondroitin sulfate, reported to treat osteoarthritis. Here we present the first multi-tissue study yielding hundreds of novel pQTLs. These results highlight the need to implement additional functional genomic approaches beyond gene expression to understand the biology of complex traits.. This first multi-tissue study will be instrumental to map signals from genome-wide association studies (GWAS) onto functional genes, to discover pathways from different proteins associated with the same pleiotropic region, and to identify drug targets for neurological diseases.

PrgmNr 1284 - Single-cell RNA-sequencing of human brain regions identifies cell types and states relevant to disease

View session detail

Author Block: **S. Gerges**[1], M. Goldman[1], M. Daly[2], S. Berretta[1], S. A. McCarroll[3]; [1]Broad Inst., Cambridge, MA, [2]Harvard Med. Sch., [3]Harvard Med. Sch., Boston, MA
Disclosure Block: **S. Gerges:** None.
Genetic association studies have identified a plethora of loci associated with brain disorders. However, the biology that mediates how genetic etiology manifests into disease remains elusive. Here, we show that single-cell RNA-sequencing (scRNA-seq) allows for unprecedented resolution into the cell types most relevant to psychiatric and neurological disease etiology as well as complex traits more broadly. We combine single-cell RNA-seq of over 120,000 cells from human choroid plexus, subventricular zone, amygdala, caudate and temporal cortex from across multiple individuals with genome-wide association study (GWAS) summary statistics and rare-variants from over 25 brain-related traits. We show that loci and genes associated with brain traits map to distinct expression signatures within diverse cell populations within the brain. We use partitioned LD score regression (LDSC) to show that schizophrenia, bipolar disorder, educational attainment and IQ are most enriched within amygdala excitatory neurons, as well as a novel cell-type population known as â□□eccentricâ□□ spiny projection neurons (eSPN) (p=1.29e-15, p=9.83e-11, p=9.22e-10 and p=3.72e-10 respectively), as well as both â□□directâ□□ (dSPN) and â□□indirectâ□□ (iSPN) populations in the patch and matrix compartments of the striatum (p We contrast this with other traits such major depressive disorder which is exclusively enriched in GABAergic neurons from the amygdala (p To our knowledge, this is the most in depth scRNA-seq survey of the amygdala, subventricular zone, cortex, caudate nucleus and choroid plexus. All regions that are heavily implicated in brain disorders, and provide an atlas for understanding the etiology of complex brain phenotypes. Altogether, we show that combining genetics with human single-cell data allows for biologically interpreting how genetics manifests within cells that contribute to disease onset.

PrgmNr 1285 - Joint transcriptome and chromatin accessibility profiling of the developing human brain cortex at the single cell level

View session detail

Author Block: **K. Zhu**, J. Fullard, J. Bendl, S. Jiang, G-C. Yuan, P. Roussos; Icahn Sch. of Med. at Mount Sinai, New York, NY
Disclosure Block: **K. Zhu:** None.
We performed multimodal profiling for gene expression and chromatin accessibility in 45,549 single nuclei isolated from the human cortex, across a broad developmental period (fetus, infancy, childhood, adolescence and adulthood). Integrative analysis of both modalities improves the *de novo* taxonomy compared to analysis using either single modality. We identified cell-type specific domains in which chromatin accessibility is highly correlated with gene expression. To better understand the regulatory mechanisms driving expression of lineage-specific genes during neurogenesis, we performed pseudotime trajectory analysis. Differentiation trajectory of neuronal populations displayed a lineage-priming pattern, suggesting that chromatin accessibility at *cis*-regulatory elements precedes the onset of gene expression. Using lineage-specific genes and peaks, we evaluated whether the genetic loci implicated in a panel of neurodegenerative and neuropsychiatric disorders map to specific cell types. Besides replicating many known associations, we found that common variants of autism are most enriched in fetal inhibitory neurons. Our results collectively suggest that single cell-derived marker genes and peaks lead to an enhanced cell type resolution of heritability analysis.

PrgmNr 1335 - Genotype first: Clinical genomics research through a reverse phenotyping approach

View session detail

Author Block: **C. Wilczewski**, J. Obasohan, L. G. Biesecker, A. E. Katz; Natl. Human Genome Res. Inst., Bethesda, MD
Disclosure Block: **C. Wilczewski:** None.
Clinical genomics research has traditionally started with a group of phenotypically similar individuals and tested whether a common genetic cause can be found. Genomic ascertainment research utilizes a novel, reverse approach. In these â□□reverse phenotypingâ□□ studies, phenotypically unselected individuals with a common genetic variant are recruited for phenotypic assessment to test whether they have similar physical characteristics. This approach holds promise to expand our understanding of the relationships between genetic variation and phenotype. The utility of this approach requires evaluation. To characterize and quantify the utility of reverse phenotyping, we reviewed 13 studies that recontacted 190 genomically ascertained participants and performed phenotypic assessment for 60 conditions. We categorized the type of research question under investigation and evaluated what results would have been obtained if complementary research methods such as electronic health records review or phenotypic ascertainment of research subjects would have been utilized instead. We evaluated 13 studies. Six tested a hypothesized novel gene-disease association. Six evaluated the phenotypic spectrum of a known association. One study performed ex vivo reverse phenotyping using patient samples obtained as a result of genomic ascertainment. In 62% of studies, testing the gene-disease association hypothesis via electronic health record review would have resulted in the opposite conclusion compared to prospective reverse phenotyping. In 92% of these studies, the genotype-phenotype hypothesis would not have been testable without prospective reverse phenotyping studies. Our conclusion is that reverse phenotyping is a valuable and effective tool to test relationships between genetic variation and phenotypes. This is especially true where there may be a subclinical disease presentation or presentation of rare phenotypes that are not typically assessed as part of routine healthcare. This work will inform how reverse phenotyping research capabilities can be utilized more broadly and efficiently to bring the reality of precision medicine closer within reach.

PrgmNr 1336 - The knowledge gap between pharmacogenomics and medical genetics: Is there a bridge in between?

View session detail

Author Block: **B. Li**[1], M. Whirl-Carrillo[2], T. E. Klein[3]; [1]Stanford Univ., Stanford, CA, [2]Stanford Univ, Palo Alto, CA, [3]Stanford Univ, Stanford, CA
Disclosure Block: **B. Li:** None.
Precision medicine faces many challenges including the gap of knowledge between medical genetics and pharmacogenomics (PGx). Medical genetics interprets the pathogenicity of genetic variants for diagnostic purposes while PGx investigates the genetic influences on drug responses. Ideally, the quality of health care would be improved from the point of disease diagnosis to drug prescribing and adherence if PGx knowledge was included. However, PGx variants are usually not reported as a secondary finding even if they are included in a clinical genetic test for diagnostic purposes despite the fact that the detection of PGx variants can provide valuable drug prescribing recommendations. One underlying reason is the lack of systematic classification of the knowledge overlap between PGx and medical genetics. Here, we cross-referenced annotations of pharmacogenes and pathogenic genes from multiple expert-curated knowledge databases, including the ClinGen curation efforts (1889 variant pathogenicity, 1220 clinical validity, and 526 clinical actionability), 440 CPIC gene-drug pairs, and the PharmGKB (4559 curated clinical annotations and 798 drug labels). We classified pharmacogenes and pathogenic genes based on the abundance of evidence supporting a geneâ□□s role in target phenotypes and the level of clinical actionability. In total, 678 pathogenic genes and 102 pharmacogenes were classified as genes with abundant supporting evidence; 25 genes were shared between them. These 25 genes were classified into three categories based on the connection between their pathogenic role and pharmacogenetic effect. (1) Pathogenic roles of the genes are related to their pharmacogenetic effects. Carriers of specific pathogenic genotypes can live without a clinical phenotype unless exposed to certain drugs, e.g., *RYR1* and *CACNA1S*. (2) Certain medications are contraindicated in individuals with the specific genetic disorders or carriers of specific genotypes. One example is the contraindication of valproic acid in patients with ornithine carbamoyltransferase deficiency or carriers of specific pathogenic *OTC* genotypes. (3) Medications are indicated for individuals with certain genetic conditions. E.g., velaglucerase alfa is a long-term therapy for type 1 Gaucher disease patients deficient of *GBA* enzyme activity. Interestingly, phenotypic variability has been observed in some of this class of genes, suggesting genetic heterogeneity. Overall, we characterized 25 pathogenic genes that warranted further investigation of genetic influences on drug efficacy and toxicity, which would collectively improve clinical care from disease diagnoses to drug prescribing.

PrgmNr 1337 - Population-based penetrance of clinical variants in individuals of diverse ancestries and ages

View session detail

Author Block: **I. Forrest**[1], K. Chaudhary[1], H. T. Vy[1], S. Bafna[2], D. M. Jordan[3], G. Rocheleau[1], R. Loos[4], J. H. Cho[1], R. Do[5]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Icahn Sch. of Med. at Mount Sinai, Jersey City, NJ, [3]Icahn Sch. of Med. at Mt Sinai, New York, NY, [4]The Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Icahn Sch. of Med., New York, NY

Disclosure Block: **I. Forrest:** None.

Introduction: A population-based assessment of disease risk associated with variants informs clinical decisions and risk stratification. We aimed to evaluate the population-based disease risk of clinical variants in known disease-predisposition genes. **Methods**: We performed a cross-sectional study of two population-based biobanks with linked exome and electronic health record data, the UK Biobank and Bio*Me* Biobank. Participants included 72,434 individuals carrying 37,772 clinical variants who were of a diverse range of ages and ancestries. The exposure was clinically impactful variants that were reported as pathogenic in ClinVar or predicted to cause a loss of protein function by bioinformatic algorithms in genes mediating disease via loss-of-function mechanism. The primary outcome was disease risk of impactful variants; risk difference (RD) between the proportion of affected carriers (penetrance) versus affected noncarriers was measured. **Results**: Among 72,434 study participants, 43,395 were from the UK Biobank (mean [SD] age, 57 [8.0] years; 45% male; 7% non-European) and 29,039 were from the Bio*Me* Biobank (mean [SD] age, 56 [16] years; 40% male; 68% non-European). Biomarker levels were significantly associated with variant penetrance, such as in *LDLR* (0.53 mg/dL increase in low-density lipoprotein cholesterol per 1% penetrance increase; $P=4x10^{-3}$), *HNF1A* (0.76 mg/dL increase in glucose per 1% penetrance increase; $P=8x10^{-3}$), and *UCP3* (0.24 kg/m^2 increase in body mass index per 1% penetrance increase; $P=6x10^{-4}$). Of 5,359 impactful variants, 4,794 (89%) conferred RD â¤0.05. Mean penetrance was 6.9% (SD=23%) for pathogenic variants and 0.85% (SD=6.4%) for benign variants reported in ClinVar. A subset of variants was ancestry-specific (460 [8.6%]), such as the *HBB* S10fs variant significantly associated with thalassemia in Asians (RD=0.99; $P=9x10^{-6}$). Variant penetrance for late onset diseases was modified by carrier age, with mean penetrance 1.8% higher in â¥70-year-old carriers compared to â¥20-year-old carriers. Penetrance of impactful variants was heterogeneous even in known disease-predisposition genes, including *BRCA1* (mean [SD] variant penetrance, 38% [45%]), *BRCA2* (38% [46%]), and *PALB2* (26% [42%]). **Conclusions**: In two populations of diverse ancestries and ages, most impactful variants were associated with a low risk of disease. Current variant interpretation categories are not concordant with disease risk. Further research of population-based penetrance is needed to refine variant interpretation and guide screening and management of individuals who carry these variants in the general population.

PrgmNr 1338 - Common-variant dysregulation of Mendelian disorder genes contributes to complex disease

View session detail

Author Block: **E. McArthur**[1], M. L. Benton[2], C. Tubbs[1], D. Ruderfer[1], N. J. Cox[3], J. A. Capra[4]; [1]Vanderbilt Univ., Nashville, TN, [2]Baylor Univ., Waco, TX, [3]Vanderbilt Univ Med Ctr., Nashville, TN, [4]Univ. of California San Francisco, San Francisco, CA

Disclosure Block: **E. McArthur:** None.

Introduction: Although Mendelian disorders and common complex traits are often framed as distinct, there is increasing evidence that they exist on a continuum. Quantifying the shared genetic underpinnings of Mendelian and complex diseases will enable translation of insights into biological mechanisms, diagnosis, and treatment for both rare and common disease. Yet, we lack a framework for evaluating how common variation near Mendelian genes contributes to related traits at a personalized level.

Hypothesis: We hypothesize that common-variant-based dysregulation of Mendelian genes will associate with risk for related complex traits in relevant tissues. For example, do individuals with common-variant-mediated downregulation of monogenic diabetes genes have an increased risk of adult-onset diabetes?

Methods: To evaluate the relationship between genetic regulation of Mendelian genes and their functional effects, we use Vanderbiltâ□□s electronic health record (EHR)-linked biobank of 90,000+ genotyped individuals (BioVU). For each individual, we predict tissue-specific regulation of Mendelian disorder genes across 49 tissues using a PrediXcan/Transcriptome-wide Association Study framework trained with transcriptome data (GTEx v8). For each Mendelian gene, we test if dysregulation is associated with clinical phenotypes in the EHR across 1865 phecodes.

Results: We comprehensively quantify how common-variant regulation of Mendelian genes influences the medical phenome. We identify specific Mendelian genes with phenotypes shared by both monogenic disruption and common-variant dysregulation. For example, disruption of *WFS1* by coding variants causes Wolfram syndrome, a form of monogenic diabetes mellitus. We find that downregulation of *WFS1* by common regulatory variants is strongly associated with risk for diabetes, demonstrating a mechanism for both complex and monogenic diabetes pathogenesis. Then, we assess which phenotypes have a shared genetic basis between common and rare variants by testing regulation-mediated associations for genes in different phenotype domains. Finally, we consider unresolved questions about rare disease. For example, we show that individuals with downregulation of genes that cause Hereditary Hemochromatosis have higher rates of thyroid disease and certain cancer subtypes in the EHR. Thus, this approach may help prioritize screening for individuals with Mendelian disease and provide new avenues for testing phenotype expansion hypotheses. Ultimately, investigation into common-variant dysregulation of Mendelian genes will further our understanding of both common complex and rare Mendelian disorders.

PrgmNr 1339 - Understanding the importance of racial diversity in the development and implementation of pharmacogenomic guidelines in kidney transplant patients receiving tacrolimus

View session detail

Author Block: **Y. Ko**[1], M. Roberts[1], A. A. Seyerle[2]; [1]Univ. of North Carolina, Chapel Hill, NC, [2]Univ. of North Carolina at Chapel Hill, Apex, NC
Disclosure Block: **Y. Ko:** None.
Tacrolimus is a commonly prescribed immunosuppressant used to prevent organ rejection. Multiple studies show *CYP3A5* genotype is associated with tacrolimus serum concentration. However, in clinical settings tacrolimus is still dosed using a trial and error method with frequent trough concentration monitoring. Our study investigated the projected clinical benefit of implementing pharmacogenomics guidelines in tacrolimus dosing among kidney transplant patients. The primary outcome was acute rejection at one month post kidney transplant. Variables impacting trough level were simulated (e.g. age, body mass index, steroid use, panel reactive antibody present, genotype, number of human leukocyte antibodies mismatch). Simulated populations were analyzed under three different dosing scenarios: weight based dosing, pharmacogenomics based dosing accounting for *CYP3A5 *1* and *3* alleles, and comprehensive pharmacogenomics based dosing accounting for *CYP3A5 *1, *3, *6,* and *7*. Under all scenarios, the percentage of populations experiencing acute rejection was slightly higher in European Americans (4.5-7.6%) compared to African Americans (3.3-6.3%), with acute rejection rates lowest under our pharmacogenomics dosing scenarios (African Americans: 3.3-5.9%, European Americans: 4.5-7.5%) compared to weight based dosing (African Americans: 5.1-6.3%, European Americans: 6.2-7.6%). African Americans achieved the highest percent of the population in the targeted trough level (8-12ng/ml) using a dose of 0.1 mg/kg/day (35.2-41.1% compared to 36.4-39.0%) whereas European Americans achieved the greatest proportion in the target range using a dose of 0.125 mg/kg/day (41.5-42.1% compared to 26.7-41.2%). Our study demonstrated that African and European Americans achieved a higher percentage of patients within target trough levels using different dosing algorithms, suggesting that maximizing the clinical benefit of precision medicine may require investigating ancestry specific dosing algorithms. Furthermore, we found African Americans had a reduced risk of acute rejection compared to European Americans, which contradicts current literature. Given that our study assigned a consistent dosing level to all participants at baseline, this suggests that the higher rates of rejection in African Americans may be explained by the higher levels of interindividual dose variability seen in African Americans based on current weight based dosing algorithms. This further emphasizes the need to achieve accurate dosing levels early in treatment, a goal which can be accomplished more effectively with the use of pharmacogenomics based dosing.

PrgmNr 1340 - Genome-wide association of structural variations with lipid metabolism traits in TOPMed GOLDN study

View session detail

Author Block: **Y. Chen**[1], M. Gao[2], H. K. Tiwari[3], D. K. Arnett[4], M. R. Irvin[1], Z. Chong[2]; [1]Univ. of Alabama at Birmingham, Birmingham, AL, [2]UAB, Birmingham, AL, [3]Univ Alabama at Birmingham, Birmingham, IN, [4]Univ Alabama Birmingham, Birmingham, IN
Disclosure Block: **Y. Chen:** None.
Structural variations (SVs) contribute substantially to the divergence between species and polymorphisms within species. However, how SVs associate with specific traits in humans is less studied. Previous genome-wide association studies (GWAS) mainly focused on smaller genetic variants, such as SNPs and indels. And the balanced accuracy and resolution of these variants were also limited due to old platforms, SNP array for example. It is very critical to understand the impact of all genetic factors on complex human diseases at the population level. The Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study from Trans-Omics for Precision Medicine (TOPMed) program has generated a wealth of genomics and clinical phenotype data to assess how genetic factors interact with environmental (diet and drug) interventions. Previous GOLDN GWAS has identified several SNPs or genes significantly associated with relevant traits such as fasting lipids, fenofibrate treatment, and so on. Here, we present a genome-wide SV association study on GOLDN dataset. SV discovery using existing genotyping is a challenging problem. We developed an SV discovery and genotyping pipeline that enables accurate downstream GWAS analysis. It includes initial SV detection by Manta and a customized SV filter based on the number of split reads and abnormal read pairs with different thresholds for different sizes of SV events. To evaluate the precision of the SV discovery pipeline, we applied the pipeline on a benchmark sample with both short-read and long-read data. With the long-read data as the ground truth, we assessed the performance of the SV discovery pipeline and validated a precision over 90% for both deletion and insertion calls. We then applied the pipeline to 959 participants in the GOLDN study to generate high-confidence SV callsets, and we merged and genotyped the SV calls using a customized script. A family-based quantitative trait association analysis with the 29,050 common (AF>0.01) SVs highlighted 39 SVs significantly (pEPS8L1 is significantly associated with several fasting lipids (Triglyceride (p=2.9E-7), VLDL (p=3.2E-7), large VLDL (p=2.5E-6), and Chylomicrons (p=3.0E-6)); a 72bp deletion on *SLC1A7* is significantly associated with Glucose level (p=1.3E-6); a 73bp deletion on *IL2RA* is significantly associated with IL-2 sR alpha, an inflammation marker (p=6.2E-7). These results suggest that these SVs may be functionally important for lipid metabolism in the Caucasian population. Future detailed functional assays of these genes by introducing SVs in vivo and/or in vitro may provide further clues.

PrgmNr 1343 - Genetic mapping across autoimmune diseases reveals shared associations and mechanisms

View session detail

Author Block: **M. R. Lincoln**[1], S. G. Chun[2,3], S. R. Sunyaev[2,3], C. Cotsapas[4]; [1]Yale Sch. of Med., New Haven, CT, [2]Harvard Med. Sch., Boston, MA, [3]Brigham & Women's Hosp., Boston, MA, [4]Yale Sch. Med., New Haven, CT

Disclosure Block: **M.R. Lincoln:** None.

Autoimmune and inflammatory diseases are heterogeneous disorders, where activation of innate and adaptive immunity, coupled with loss of self-tolerance leads to target tissue destruction. Genome-wide association studies (GWAS) have identified hundreds of susceptibility loci, including many loci that contain risk alleles for multiple diseases. It has been difficult to assign specific molecular mechanisms to individual risk loci, particularly as genetic resolution is limited by available cohort size. In this study, we show that many overlapping genetic associations in fact share common risk alleles. By jointly analyzing shared associations, we improve fine-mapping resolution and identify specific downstream mechanisms for multiple autoimmune diseases.

Applying LD score regression to 19 autoimmune disease GWAS, we first demonstrate substantial genome-wide shared heritability among autoimmune and inflammatory diseases. We observe a complex pattern of sharing, with strong correlations between individual pairs of diseases (e.g. among IBD subtypes, 0.63 â¤ r_g^2 â¤ 0.92) and weaker correlations among others (e.g. MS and RA, r_g^2 = 0.15). The pattern suggests multiple shared mechanisms that differ between pairs of diseases.

We then examine 224 instances where genetic associations to multiple diseases occur in the same region, using a collection of 129,058 cases and controls genotyped on the ImmunoChip. Using joint likelihoods, we show that 41.5% of these observed associations are due to pleiotropic variants, with the remainder being due to different alleles. We use meta-analysis to combine cases and controls across diseases, increasing fine-mapping resolution at shared loci two-fold on average. Mean 95% credible interval size decreases from 37.2 SNPs for individual diseases, to 17.3 SNPs upon meta-analysis across diseases.

With improved resolution, we identify new disease risk variants that alter gene expression in immune cell subtypes. Again using joint likelihoods, we identify 21 new disease:eQTL overlaps at shared loci (from 139 in individual diseases) in the BLUEPRINT dataset. Specificity is increased, with 11 prior disease:eQTL overlaps no longer significant in meta-analysis. Gains occur primarily in cases where meta-analysis reduces credible interval size, indicating that improved resolution drives these new associations.

Our results show that a large proportion of overlapping associations among autoimmune diseases are in fact shared, and that joint analysis of these uncovers shared mechanisms. Our approach can be applied to any set of traits, and is particularly valuable as existing sample collections become depleted.

PrgmNr 1344 - Systematic discovery of gene-environment interactions for metabolic serum biomarkers

View session detail

Author Block: **K. Westerman**[1], F. Giulianini[2], J. C. Florez[1], H. Chen[3], D. I. Chasman[4], M. S. Udler[5], A. K. Manning[1], J. B. Cole[6]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [4]Brigham & Women's Hosp, Boston, MA, [5]Broad Inst., Cambridge, MA, [6]The Broad Inst. of MIT and Harvard, Cambridge, MA
Disclosure Block: **K. Westerman:** None.

Gene-environment interactions (GEIs) inform precision genomic medicine by describing how environmental exposures modify the effects of genetic variants on disease risk. Comprehensive mapping of GEIs across all genetic variants, exposures, and outcomes is challenged by the massive search space and accompanying low statistical power. Variance-quantitative trait loci (vQTLs), or genetic variants that associate with differential variance (rather than mean) of an outcome, are enriched for underlying GEIs and can help prioritize variants for subsequent GEI testing. Here, we develop a catalog of GEIs affecting metabolic biomarkers using an initial genome-wide vQTL scan followed by systematic GEI testing across thousands of exposures.

A two-stage analysis was conducted in 355,790 unrelated participants in the UK Biobank, stratifying across four ancestries followed by meta-analysis. In stage one, genome-wide vQTL scans were performed using Leveneâ□□s test for differential phenotypic variance across genotypes for 20 serum metabolic biomarkers (e.g., lipids, lipoproteins, and glycemic measures) identifying 182 independent locus-biomarker relationships. While 90.5% of these variants were also known GWAS mean-effect loci, they represent only a small fraction (3%) of such loci, meaning that vQTL prioritization substantially reduced the GEI search space. 26 vQTLs (e.g., *APOE*) were shared across multiple biomarkers, and over 40% replicated nominally in the much smaller Womenâ□□s Genome Health Study (N=23,294), demonstrating robustness in our results.

In stage two, GEIs underlying these vQTLs were revealed by testing over 2300 socioeconomic, lifestyle, and clinical exposures. Across all pairwise combinations of exposures and vQTLs, 888 GEIs passed a stringent significance threshold. Many involved anthropometric exposures, where preliminary clustering analysis revealed distinct clusters of variants interacting with either fat-free mass- or adiposity-related traits. Other interactions had implications for lifestyle choices: alcohol intake was associated with an increase in ALT, a liver marker, specifically in carriers of two missense alleles at rs1229984 in the alcohol-processing *ADH1B* gene (p_{int}=8.310^{-13}).

Our vQTL scan prioritized a subset of well-known GWAS loci that affect metabolic biomarker variance, while the following systematic GEI analysis revealed hundreds of novel interactions and pinpointed specific participating exposures. Our catalog of vQTLs and GEIs, which is publicly available on the Common Metabolic Disease Knowledge Portal, provides a knowledge base for genome-centered precision approaches to metabolic health.

PrgmNr 1345 - Global Biobank Meta-analysis Initiative: power genetic discovery for human diseases with > 2.6 million samples across diverse ancestries

View session detail

Author Block: **W. Zhou**[1], M. Kanai[2], A. R. Martin[3], K-H. Wu[4], J. Karjalainen[5], M. Kurki[6], S. Chapman[2], C. J. Willer[4], B. M. Neale[3], M. J. Daly[3], Global Biobank Meta-analysis Initiative; [1]Massachusetts general Hosp., Broad Inst., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Massachusetts Gen. Hosp., Boston, MA, [4]Univ. of Michigan, Ann Arbor, MI, [5]Boston, MA, [6]

Disclosure Block: **W. Zhou:** None.

The Global Biobank Meta-analysis Initiative (GBMI) brings together large-scale genomic studies by biobanks to foster collaboration in understanding the genetic basis of human disease. 21 biobanks have joined the initiative, totaling more than 2.6 million individuals with matched genetic data and electronic health records. In the flagship project, we meta-analyzed up to 18 biobanks on 14 endpoints of common interest including common or well-studied diseases, such as asthma, less prevalent or understudied diseases, such as gout and acute appendicitis, and the relevant procedure endpoint, appendectomy. GBMI incorporates diverse ancestries in genetic studies by including biobank samples across 6 main populations: 341,000 East Asians, 31,000 South and Central Asians, 33,000 Africans, 18,000 Admixed Americans, 156,000 Finnish, and 1.2 million Non-Finnish Europeans. Analysis pipelines for conducting GWAS with harmonized phenotype definitions in individual biobanks and cloud-based infrastructure for sharing data and performing meta-analyses have been well established. All-biobank meta-analyses of 14 endpoints have successfully replicated 271 previously reported loci and identified 237 potentially novel loci. The effect sizes of 33 loci are heterogeneous across biobanks, 35 are heterogeneous across ancestry, and 47 are different in males and females. By evaluating the effect sizes and association p-values for index variants in previous studies and employing genetic correlation, we have validated the integration of genetic studies across biobanks despite the different underlying raw data, such as phenotype ascertainments (ICD vs. self-reported) and sample recruiting strategies (hospital vs. general population) and demonstrated the improved association power given the increased sample sizes. Out of the 10 loci identified for acute appendicitis by meta-analyzing 10 biobanks, 3 were also significant for appendectomy even though the sample size was 3 times lower, suggesting that the procedure endpoints can be useful in biobank-based genetic studies. Besides gleaning deeper biological implications for human diseases with extensive in-silico analyses, multiple groups within GBMI are working on demonstrating the value of such collaborative efforts, e.g. increased prediction accuracy of polygenic risk scores. GBMI is also addressing challenges commonly faced in biobank-based genetic studies such as establishing best practices to prioritize functional genes and variants which currently remains unclear. This flagship project provides considerations and guidelines for future work in biobank meta-analyses.

PrgmNr 1346 - Phenome-wide HLA association landscape of 235,000 Finnish biobank participants

View session detail

Author Block: **J. Partanen**, S. Koskela, K. Hyvarinen, FinnGen consortium, J. Ritari; FRC Blood Service, Helsinki, Finland
Disclosure Block: **J. Partanen:** None.
The human leukocyte antigen (HLA) system is the single most important genetic susceptibility factor for many autoimmune diseases and immunological traits. However, in a range of clinical phenotypes the impact of HLA alleles or their combinations on the disease risk are not comprehensively understood. For systematic population-level analysis of HLA-phenotype associations we imputed the alleles of classical HLA genes in a discovery cohort of 146,630 and replication cohort of 89,340 Finns of whom SNP genotype data and 3,355 disease phenotypes were available as part of the FinnGen project. In total, 3,649 statistically significant single HLA allele associations in 368 phenotypes were found in both cohorts. In addition to known susceptibility alleles, we discovered a number of previously poorly-established HLA associations. *DRB1*04:01-DQB1*03:02*, a frequent high-risk haplotype for many autoimmune diseases, was also independently associated with infectious diseases. Conditional analyses to distinguish protective effects from nonpredisposition showed that in 21 disease categories the effect of the high-risk allele was significantly modified by the other allele of the same gene. Furthermore, in many immunological diseases the strength of the top risk allele was significantly modified by an allele of another HLA gene. The results highlight the complex structure of HLA-disease associations and suggest that the entire HLA composition should be considered in genetic risk estimation and functional studies. Shared HLA alleles in autoimmune and infectious diseases support a link between environmental exposure and immunogenetics in these diseases.

PrgmNr 1347 - Global discovery of genetic risk variant allelic enhancer activity in allergic eosinophilic esophagitis and atopic dermatitis

View session detail

Author Block: **X. Lu**, X. Chen, Y. Huang, C. Forney, M. Granitto, A. Diouf, O. Donmez, S. Parameswaran, M. T. Weirauch, L. C. Kottyan; Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH
Disclosure Block: **X. Lu:** None.

Eosinophilic Esophagitis (EoE) is a chronic antigen-induced allergic inflammatory disease characterized by marked esophageal eosinophilia and esophageal dysfunction. While EoE is rare (1:1000), around 20% of patients with EoE will also develop atopic dermatitis (AD), which is one of the most prevalent inflammatory skin diseases. EoE and AD are characterized by Th2-associated inflammation marked by the production of Th2 cytokines, such as IL-4/IL-13, and the recruitment of Th2 cells into the inflamed tissues. IL-13 is involved in EoE and AD pathogenesis in *in vivo* models. Genome-wide association studies have identified disease-specific and common genetic predisposing factors to EoE and AD, with 531 genetic variants at 41 risk loci for EoE and 3,006 variants at 122 loci for AD. The vast majority of these variants are non-coding, suggested they likely act by impacting gene regulatory mechanisms. To systematically screen these variants for their effects on gene regulation, we constructed a massively parallel reporter assay (MPRA) library comprising 14,192 DNA oligonucleotides containing the genomic context around every allele of these 3,540 variants. Transfection into TE7 (esophagus) and HaCaT (keratinocyte) cells with or without IL13 stimulation revealed 577 variants with enhancer activity, with 138 variants showing genotype-dependent (allelic) enhancer activity. All 138 variants can be nominated as putative causal variants for these risk loci. Surprisingly, IL-13 stimulation doesnâ□□t have a significant effect on variants with enhancer activity, suggesting that IL-13 might not play an important role in gene regulatory mechanisms at EoE or AD risk loci in these cell types. These enhancer variants display strong disease specificity, with the predicted target genes for EoE and AD enriching for different disease-relevant pathways. We identified cell-type differences, with more than half of the variants not shared between two cells, indicating cell-type specific regulation of genetic variant activity. Transcription factor binding site motif enrichment analyses in the regions containing these variants with (allelic) enhancer activity reveals strong enrichment for a set of nuclear receptors in both diseases, including RARÎ³, RXRÎ± and LXRÎ². Collectively, these results reveal widespread, largely disparate allelic activity for these diseases, and reveal a potential common disease mechanism involving nuclear receptors regulating the activity of genetic variants for disease susceptibility.

PrgmNr 1348 - A genome-wide survey of recurrent repeat expansions in cancer genomes

View session detail

Author Block: **G. S. Erwin**; Stanford Univ., Palo Alto, CA
Disclosure Block: **G.S. Erwin:** None.
Expansion of a single repetitive DNA sequence, termed a tandem repeat (TR), is known to cause more than 30 rare but devastating diseases, including myotonic dystrophy and Fragile X syndrome. TR sequences can be short (150 bp), and indeed repeat expansions that create long TRs are the loci that are usually pathogenic. However, uniquely mapping repeat expansions with short-read DNA sequencing is difficult because TRs are ubiquitous in the genome. A failure to understand genetic variation in long TRs would make it impossible to capitalize on this information to develop new TR-targeting therapeutics, which I previously showed can rescue expression of genes dysregulated in disease (Erwin et al., Science 2017). Despite their importance to monogenic disease, the frequency and function of repeat expansions that create long TRs are unknown in complex genetic diseases such as cancer.We analyzed >2,500 cancer genomes from the PCAWG dataset with Expansion Hunter De Novo, a new bioinformatic tool that has been validated with both known and novel repeat expansions. We identified 79 recurrent repeat expansions (rREs) in 6 human cancers. Several repeat expansions were validated in an independent cohort of samples. In 5% of all cancers, expansions and contractions accumulate in short (UGT2B7, proximal to an enhancer. This expansion is correlated with a decrease in *UGT2B7* expression, and low expression of *UGT2B7* is associated with poor survival outcomes. Our data provide evidence that rREs are an important but overlooked source of genetic variation in human cancer.

PrgmNr 1351 - Brain-specific succinyl-CoA synthetase deficiency leads to increased protein succinylation, abnormal neuronal metabolism, and altered chromatin accessibility

View session detail

Author Block: **M. Anderson**[1], P. Gillespie[1], X. Chu[1], H. Gao[1], M. Marcus[1], Y. Liu[1,2], Y. Wang[1], B. H. Graham[1]; [1]Dept. of Med. and Molecular Genetics, Indiana Univ Sch. of Med., Indianapolis, IN, [2]Ctr. for Computational Biology & Bioinformatics, Indiana Univ Sch. of Med., Indianapolis, IN
Disclosure Block: **M. Anderson:** None.
Succinyl-CoA Synthetase (SCS) catalyzes the reversible conversion of succinyl-CoA to succinate in the Krebâ□□s cycle. In humans, pathogenic variants in SCS subunits are associated with mitochondrial encephalomyopathy with tissue-specific mitochondrial DNA (mtDNA) depletion. In eukaryotes, SCS is a heterodimer with a single catalytic alpha subunit, Suclg1, and two isoforms of its beta-subunit (Sucla2 or Suclg2). Using a CRISPR-Cas9 generated floxed allele of *Sucla2* and the forebrain-specific *CamKIIÎ±*-Cre, forebrain-specific knock-out of *Sucla2* has been achieved, resulting in the first *in vivo* adult model of homozygous SCS deficiency. Molecular validation via RT-qPCR demonstrates significant reduction of *Sucla2* gene expression in mutant cortex. However, western blot analysis indicates that forebrain-specific knockout of SUCLA2 results in reduced protein expression of both SUCLG1 and SUCLG2, suggesting significant post-transcriptional down-regulation of the entire SCS complex. Notably, this model does not exhibit mtDNA depletion, allowing investigation of brain-specific pathogenic mechanisms of SCS deficiency independent of metabolic defects caused by mtDNA depletion. Enzyme activity analysis revealed significantly reduced function of SCS, predicted to result in increased cellular succinyl-CoA. Accumulation of cellular succinyl-CoA is further corroborated by observed metabolic perturbations, including elevated C4-DC acyl-carnitines and methylmalonic acid (MMA). Succinyl-CoA is the substrate for protein lysine-succinylation, a post-translational modification shown to have impact on multiple biological processes, including many metabolic pathways and histone modification. Global protein succinylation is significantly increased in the cortex of the mutant model, and ATAC-seq results demonstrate significant alterations in chromatin accessibility. Ongoing studies include proteomics and transcriptomics to further elucidate the consequences of SCS deficiency in neurons. This unique model promises to provide new insight into the function of SCS and lysine succinylation on brain metabolism, which will have a significant impact on therapeutic research of metabolic and mitochondrial disorders.

PrgmNr 1352 - Aberrant posttranslational modifications contribute to MMA pathophysiology and identify new targets for therapy

View session detail

Author Block: **P. E. Head**[1], S. Myung[1], Y. Chen[1], D. Romero[1], J. Schneller[1], S. McCoy[1], I. Manoli[2], M. Gucek[1], C. P. Venditti[3]; [1]NIH, Bethesda, MD, [2]NIH, Bethesda, MD, [3]NHGRI (NIH), Bethesda, MD
Disclosure Block: **P.E. Head:** None.
Organic acidemias, such as methylmalonic acidemia (MMA), are a group of inborn errors of metabolism that typically arise from defects in the catabolism of amino- and fatty acids. The accretion of acylic acid species is postulated to cause intracellular toxicity and underlie the dysregulation of multiple intermediary pathways seen in the patients, such as the urea cycle and glycine cleavage system. Here, we explore an alternative pathophysiological consequence of impaired acyl-CoA metabolism: the accumulation of aberrant posttranslational modifications (PTMs) on enzymes in critical intracellular pathways. Using an MMA mouse model that recapitulates MMA hepatic mitochondriopathy (*Mmut*$^{-/-}$;*Tg*$^{INS\text{-}MCK\text{-}Mmut}$), we surveyed PTMs in hepatic extracts with acyl-lysine antibodies and discovered widespread hyper-acylation. Next, we performed mass spectrometry to characterize the PTM proteome. Excessive acylation of enzymes involved in glutathione, urea, arginine, tryptophan, valine, isoleucine, methionine, threonine, and fatty acid metabolism were detected in the MMA mice, and validated via immunoprecipitation analysis. We extended our analyses to other MMA perturbed pathways, including the glycine cleavage system, which we found to be aberrantly modified in liver extracts from MMA patients. These discoveries provide the mechanism underlying both hyperammonemia and ketotic hyperglycinemia, secondary perturbations of MMA. We have since identified the sirtuin deacylase capable of removing MMA-related PTMs and using rational mutagenesis, we created a â□□SuperSIRTâ□□ that was resistant to acylation-dependent inhibition, validated activity in vitro, cloned it behind a liver specific promoter (LSP), and packaged it within an AAV8 capsid. The resulting AAV8 LSP SuperSIRT or an AAV8 LSP EGFP control were then systemically delivered to *Mmut*$^{-/-}$;*Tg*$^{INS\text{-}MCK\text{-}Mmut}$ and *Mmut*$^{+/-}$;*Tg*$^{INS\text{-}MCK\text{-}Mmut}$ mice, at a dose of 1e13 GC/kg, and followed by biochemical and enzymatic analyses. Blood ammonia levels were significantly reduced in the treated mutant mice compared to the EGFP and untreated mutant control groups, while plasma methylmalonic acid levels remained unchanged. In hepatic extracts from the SuperSIRT treated *Mmut*$^{-/-}$;*Tg*$^{INS\text{-}MCK\text{-}Mmut}$ mice, aberrant acylation on key protein targets in the urea cycle and glycine cleavage system was reversed compared to GFP treated controls. In summary, our studies have identified a new PTM axis in patients and mice with MMA, which has allowed the development of a SuperSIRT gene therapy that could be used to treat all forms of MMA and might be extended to other disorders where aberrant acylation plays a role in disease pathophysiology.

PrgmNr 1353 - The gnomAD database of mitochondrial DNA variation across 56,434 individuals

View session detail

Author Block: **S. Calvo**[1,2], K. M. Laricchia[3], N. J. Lake[4], N. A. Watts[1], M. Shand[1], A. Haessly[1], L. Gauthier[1], D. Benjamin[1], E. Banks[1], K. Garimella[1], Genome Aggregation Database Consortium, D. G. MacArthur[5], G. Tiao[1], M. Lek[6], V. K. Mootha[7]; [1]Broad Inst., Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Yale Univ., New Haven, CT, [5]Garvan Inst., Darlinghurst, Australia, [6]Dept. of Genetics, Yale Univ., New Haven, CT, [7]Howard Hughes Med. Inst. and Massachusetts Gen. Hosp., Boston, ME
Disclosure Block: **S. Calvo:** None.
Databases of allele frequency are extremely helpful for evaluating clinical variants of uncertain significance, however until now genetic databases such as the Genome Aggregation Database (gnomAD) have ignored the mitochondrial genome (mtDNA). We have developed a pipeline to call mtDNA variants that addresses three technical challenges: (i) heteroplasmic variants present in any fraction of mtDNA molecules, (ii) a circular genome, and (iii) mis-alignment of nuclear sequences of mitochondrial origin (NUMTs). We apply this pipeline to 56,434 whole genome sequences in the gnomAD v3.1 database that includes individuals of European (58%), African (25%), Latino (10%), and Asian (5%) ancestry. We find that the number of mtDNA copies per cell varies across gnomAD cohorts and directly relates to accuracy in distinguishing heteroplasmic variants from NUMT-derived false positives, and thus we conservatively report only variants with heteroplasmy above 10%. Our gnomAD v3.1 release contains population frequencies for 10,850 unique mtDNA variants, and includes both homoplasmic (98%) and heteroplasmic (2%) variant calls. We observe that 1/250 individuals carry a pathogenic mtDNA variant with heteroplasmy >10%. The gnomAD database of population allele frequencies will be useful for evaluating clinical mtDNA variants of unknown significance, and will become increasingly useful as the database grows. The mtDNA variants are publicly available within gnomAD v3.1 (gnomad.broadinstitute.org).

PrgmNr 1354 - Clinician lead analysis of WGS data identifies missed diagnoses in suspected primary mitochondrial disease

View session detail

Author Block: **W. Macken**, C. McKittrick, C. Pizzamiglio, E. Bugiardini, M. Falabella, R. Quinlivan, H. Holden, M. Hanna, Genomics England Research Consortium, J. Vandrovcova, R. D. S. Pitceathly; UCL Inst. of Neurology, London, United Kingdom
Disclosure Block: **W. Macken:** Grant/Contracted Research Support (External); Medical Research Council.
Purpose: Clinical WGS promises improved diagnostic rates in rare disease, however, high-throughput approaches in clinical laboratories may overlook diagnoses with nuanced presentations. A broader approach is especially needed in conditions like Primary Mitochondrial Diseases (PMDs), where phenotypes are non-specific. We sought to assess whether a clinically focused review could increase the diagnostic rate of these complex patients. **Methods used:** One hundred and two probands attending a Specialised Service for PMDs underwent WGS (Illumina TruSeq, HiSeq 2500) via the 100,000 Genomes project following exclusion of common molecular causes of PMD. Standard virtual panel analysis was undertaken initially. The phenotype of undiagnosed patients was then scrutinised by a clinician, their data was re-annotated by a clinical bioinformation, and subsequently manually reviewed by a clinician on a clinical interpretation platform. Re-analysis of data included expansion of gene panels applied if appropriate, and analysis for cryptic in trans variants in strong candidate recessive genes, assessment for CNVs, and interrogation of mtDNA using a somatic variant caller. **Key Results:** The standard bioinformatic pipeline with clinical scientist-driven interpretation yielded diagnoses in 18/102 (17.6%). Expanded analysis resulted in a diagnostic uplift in 17/102 (16.7%) of cases; heteroplasmic mtDNA variants were identified in 3/102 patients; nDNA diagnoses were confirmed in 14/102 patients; separately, suspected pathogenic variants in novel genes were identified in 3 further patients. Reasons for missed diagnoses included limited use of virtual gene panels, interpretation problems with complex family histories, missed mosaic variants, missed heteroplasmic variants and missed variants in new morbid genes. Only 40% of all diagnoses were true PMDs, emphasising the need to include a broad differential when diagnosing these patients. Our approach close to doubled the diagnostic rate from 17.6% to 13.3%. To harness the full power of WGS in clinical services, high throughput analysis by clinical scientists should be include in depth phenotype-focused analysis by a genomic medicine clinician. This research was made possible through access to the data and findings generated by the 100,000 Genomes Project; http://www.genomicsengland.co.uk.

PrgmNr 1355 - Regulating *PCCA* gene expression by splice switching antisense oligonucleotide-mediated modulation of pseudoexon splicing to rescue propionyl CoA carboxylase enzyme activity in propionic acidemia

View session detail

Author Block: **U. Petersen**[1], M. Dembic[1], A. MartÃnez-Pizarro[2], L. R. Desviat[2], B. S. Andresen[1]; [1]Dept. of Biochemistry and Molecular Biology, Univ. of Southern Denmark, Odense M, Denmark, [2]Centro de BiologÃa Molecular Severo Ochoa, Departamento BiologÃa Molecular, Univ. AutÃ³noma de Madrid, Madrid, Spain
Disclosure Block: **U. Petersen:** None.
The deep intronic sequence variation, c.1285-1416A>G, activates an 84 bp pseudoexon from intron 14 in the *PCCA* gene by changing the activity of a splicing regulatory element within the pseudoexon sequence. The *PCCA* pseudoexon is activated for increased inclusion in mRNA, however, relatively high inclusion levels can be observed in normal cells as well. Inclusion of the pseudoexon introduces an in-frame premature termination codon, which targets the pseudoexon-containing mRNA for degradation by nonsense-mediated mRNA decay. Activation of the *PCCA* pseudoexon causes deficiency of propionyl-CoA carboxylase (PCC) and propionic acidemia, a rare autosomal recessive metabolic disorder with metabolic decompensation and multiorgan complications.
We and others have shown that splice-switching antisense oligonucleotides (SSO) can be employed to inhibit inclusion of the *PCCA* pseudoexon. We have used different minigenes to study the underlying molecular defect in detail and to identify efficient SSOs. We demonstrate that SSO-mediated skipping of the *PCCA* pseudoexon increases PCC enzyme activity in patient fibroblasts homozygous for *PCCA* c.1285-1416A>G, in patient fibroblasts with *PCCA* missense variations with residual PCC activity, and also in control fibroblasts with wildtype *PCCA*. We generated a HepG2 cell line with the *PCCA* c.1285-1416A>G variation, using the CRISPR/Cas12a system, providing us with a stable isogenic model of the pseudoexon activation event. SSO-mediated skipping of the *PCCA* pseudoexon restores normal splicing patterns and expression of wildtype *PCCA*. Interestingly, it increases both PCCA and PCCB protein levels and increases PCC activity in HepG2 *PCCA* c.1285-1416A>G and in HepG2 *PCCA* wildtype cell lines. This suggests that the *PCCA* pseudoexon can be exploited as a gene-regulatory switch and that SSO-mediated modulation of *PCCA* splicing patterns can be employed in treatment of propionic acidemia, both in patients with the *PCCA* pseudoexon activating variation and in patients with pathogenic missense variations with residual PCC activity. Moreover, we show that the SSO-mediated increase of PCCA protein amounts rescues unstable overproduction of PCCB from degradation, increasing the production of PCC heterododecamers and thereby increasing the PCC enzyme activity. This suggests that SSO-mediated skipping of the *PCCA* pseudoexon could potentially increase PCC activity also in cells with pathogenic *PCCB* missense variations with residual PCC activity.

PrgmNr 1356 - Multi-omics and functional studies in two cellular models of Barth Syndrome reveal novel mitochondrial defects and suggest new therapeutic approaches

View session detail

Author Block: **O. Sniezek**[1], A. Anzmann[2], N. Senoo[3], E. Tampakakis[4], S. Claypool[3], H. Vernon[1]; [1]Johns Hopkins Sch. of Med., Dept. of Genetic Med., Baltimore, MD, [2]Johns Hopkins Univ. Sch. of Med., Dept. of Genetic Med., Baltimore, MD, [3]Johns Hopkins Sch. of Med., Dept. of Physiology, Baltimore, MD, [4]Johns Hopkins Sch. of Med., Dept. of Med.: Cardiology, Baltimore, MD
Disclosure Block: **O. Sniezek:** None.
Barth Syndrome (BTHS) is a rare X-linked disorder caused by defects in the gene *TAFAZZIN*, resulting in defective cardiolipin synthesis. Clinically, BTHS is characterized by cardiomyopathy, neutropenia, skeletal myopathy, and growth defects, and has a high morbidity and mortality. Currently, there are no approved disease-specific treatments. In order to investigate the cellular pathology in BTHS, to understand tissue-specific disease effects, and to identify new areas of potential therapeutic intervention, we developed two CRISPR-edited cell lines: *TAFAZZIN*-null HEK293 cells, with which to perform broad-based discovery experiments, and *TAFAZZIN*-null iPSCs, which were differentiated into disease-relevant cell types to study tissue-specific disease effects. Proteomics, lipidomics, and metabolomics in *TAFAZZIN*-null HEK293 cells revealed diverse mitochondrial abnormalities, including abnormal mitochondrial complex I assembly, expression and function, and upregulation of PDK2, an inhibitor of pyruvate dehydrogenase. When we inhibited PDK2 function, we observed increased expression of proteins associated with mitochondrial stress. Thus, upregulated PDK2 expression may serve as a protective mechanism in *TAFAZZIN*-deficient cells and inhibition of PDK2, a therapeutic approach for other mitochondrial disorders, may be counterproductive in Barth Syndrome. When we targeted early steps in cardiolipin metabolism and cardiolipin stability with bromoenol lactone and elamipretide, respectively, we were able to remediate Complex I dysfunction. Thus, targeting cardiolipin metabolism, both proximally and distally to the primary molecular defect, are promising directions for therapeutic targeting. We next explored tissue-specific manifestations of Barth Syndrome via lipidomics and RNA-seq in iPSC-derived *TAFAZZIN*-deficient and wild-type cardiomyocytes and neurons. These studies revealed extensive disturbances in cellular lipid content and mitochondrial gene expression between both WT and *TAFAZZIN* null cells, and between *TAFAZZIN*-null cardiomyocytes and neurons. Oxygen consumption studies showed defective maximal respiratory capacity in both *TAFAZZIN*-null cardiomyocytes and neurons compared to WT. We are currently targeting specific fatty acid pools and early steps in cardiolipin metabolism in the differentiated *TAFAZZIN*-deficient cells with the goal of remediating cellular lipid, mitochondrial gene expression, and oxygen consumption abnormalities.

PrgmNr 1359 - Systematic functional interrogation of the Lynch Syndrome gene *MLH1* by deep mutational scanning

View session detail

Author Block: **X. Jia**[1], B. B. Burugula[1], S. Jayakody[1], S. L. Hemker[1], C. Smith[2], A. Scott[1], J. O. Kitzman[1,2]; [1]Dept. of Human Genetics, Univ. of Michigan, Ann Arbor, MI, [2]Dept. of Computational Med. and Bioinformatics, Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **X. Jia:** None.
Variant interpretation poses a key barrier to the clinical utility of genetic testing. Even intensively studied genes, such as the DNA mismatch repair factors that underlie Lynch Syndrome, have a heavy burden of variants of uncertain significance (VUS). Missense variants are particularly challenging to interpret, and ~87% of the Lynch Syndrome gene missense variants listed in the ClinVar database remain unclassified. This VUS burden severely impedes diagnosis, prevention and intervention for this highly prevalent (~1:300) and actionable familial cancer syndrome. To systematically generate functional evidence that can assist clinical variant interpretation, we performed a deep mutational scan of *MLH1,* one of four key genes underlying Lynch Syndrome. We used saturation single-codon mutagenesis to generate libraries covering 99.6% of all *MLH1* missense mutations (N=14,364). These libraries were stably knocked in to *MLH1*-deficient human cell lines, and the resulting population of *MLH1* missense mutant cells was subjected to functional selection using 6-thioguanine treatment, an established assay for mismatch repair gene function. We deeply sequenced these *MLH1* libraries before and after selection, to generate loss-of-function (LoF) scores for >14,300 *MLH1* missense variants. These LoF scores were bimodally distributed, with 22% (n=3,139) of all missense variants having a damaging effect on MLH1 function. Overlaying these LoF scores onto MLH1 protein structural domains, we observe damaging variants are strongly enriched in the N-terminal ATPase domain (in which 26.8% of missense variants are LoF) and C-terminal dimerization domain (24.6%). By contrast, the internal, intrinsically disordered linker domain is highly tolerant to single missense mutations (only 7.9% of variants are LoF). These scores showed strong concordance (AUC ROC=0.995) with expert reviewed classifications from ClinVar (N=103 missense variants), suggesting that they are highly predictive of pathogenicity and can extrapolate to unclassified and newly observed variants. This map, together with a recent scan of another key Lynch gene *MSH2* from our group (Jia et al, *AJHG*, 2021), demonstrates the promise of systematic functional characterization for enabling prospective early detection and intervention in Lynch Syndrome.

PrgmNr 1360 - Comprehensive cell-type classification of tumor and normal cells from single cell RNA sequencing in pan cancer settings

View session detail

Author Block: **I. Nofech-Mozes**[1], S. Abelson[1], P. Awadalla[2]; [1]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [2]Ontario Inst Cancer Res., Toronto, ON, Canada
Disclosure Block: **I. Nofech-Mozes:** None.
Single-cell RNA sequencing (scRNA-seq) has gained popularity due to its ability to study the transcriptome at a cellular level, where populations of cells are annotated based on the expression of marker genes. Thus, scRNA-seq provides a tool to gain cell-specific, subtle insights on cancer biology. However, precise annotation of cell type remains a challenge, hindering the efficiency and accuracy of data interpretation. A number of existing tools for cell-type annotation have been developed in an attempt to improve resolution and reproducibility, yet their performance is reduced when the reference dataset contains a large number of cell types from multiple tissue types, subclasses of similar cell types, or malignant cells. The large degree of interpatient malignant cell heterogeneity often leads to reduced accuracy when classifying cancer cells as most methods rely on correlations to a reference from a different source. Given the challenges in the annotation of scRNA-seq data of cancer and its high impact for elucidating mechanisms associated with tumor heterogeneity, pathogenesis, and treatment, we developed a comprehensive, hierarchically organized, multi-layered classifier spanning diverse malignant and normal cells of the tumor microenvironment. We found that performance improves when each layer focuses on a smaller number of classes and each cell sequentially moves down a series of classifiers with increased cell type resolution. When applied to 12 external validation datasets of solid tumor biopsies spanning diverse cancer types, the classifier accurately annotated the tissue of origin of malignant cells, and relevant subtypes of stromal and blood cells, with average specificity, sensitivity, and F1 scores of 0.99, 0.89 and 0.93 for cancer cells. Using confidence thresholds at each layer, the classifier abstains from classifying ambiguous cells. When applying CHETAH, an existing hierarchical classifier provided with the same reference, the majority of cancer cells are unclassified. Moreover, we validated our classifierâ□□s capability to correctly identify the tissue of origin in scRNA-seq data derived from brain metastasis of lung adenocarcinoma. This analysis demonstrates the classifierâ□□s power to uncover the tissue of origin in cancers of unknown primary. This study provides a flexible model for the annotation of cells comprising the tumor microenvironment in pan cancer settings, while existing methods require tissue-specific references for every cancer type. Our classifier provides a powerful method for investigating intercellular communication pathways between tumor cells and non-malignant cells of the tumor microenvironment.

PrgmNr 1361 - Deep transcriptomic profiling with quantitative measures identifies high-risk multiple myeloma

View session detail

Author Block: **R. Griffin Waller**[1], B. Avery[2], M. J. Madsen[2], H. A. Hanson[2], N. J. Camp[2]; [1]Mayo Clinic, Rochester, MN, [2]Huntsman Cancer Inst. and Univ. of Utah Sch. of Med., Salt Lake City, UT
Disclosure Block: **R. Griffin Waller:** None.

Transcriptomic studies can help decipher tumor heterogeneity and identify patient risk groups. SPECTRA is an approach to describe variation in bulk transcriptome with unsupervised quantitative variables. Spectra variables represent a deep dive into the transcriptome, including both large and small sources of variance, and can be used to model with any clinical, demographic, or biological endpoint. We applied the SPECTRA approach to multiple myeloma (MM), the second most common blood cancer. Using RNA sequencing from malignant cells, we derived 39 unsupervised spectra in 767 patients in the MMRF CoMMpass study. The 39 spectra capture 65% of the variation in patientsâ transcriptomes. In multivariate Cox regression for overall survival (OS, 179 events), 9 spectra were selected as predictor variables. Bootstrap internal validation was used to adjust for over fitting. The model attained an adjusted C-index (C_{adj}) = 0.66 (0.65-0.74). Gaussian mixture modeling was used to determine high and low-risk patients. High-risk patients (n = 88, 58 events) had median survival of 1.7 years. Low-risk patients (n = 679, 121 events) did not reach median survival after 6.5 years. The OS adjusted hazard ratio (HR_{adj}) = 4.27 (2.31-12.69) for spectra high/low risk patients. We compared this to the most widely adopted transcriptomic risk score for survival, UAMS, a score based on 70 differentially expressed genes (k-means defines high/low risk-groups). In the CoMMpass data, the OS HR_{adj} = 3.89 (2.53-5.45) for UAMS high/low risk patients. To understand the overlapping and distinct signals of the two scores and clinical risk factors we used added predictive value (APV). In multivariate Cox regression for OS, including revised internal staging system (R-ISS) and age at diagnosis, spectra risk-score was highly significant ($p = 2.36x10^{-20}$) with APV = 0.612. Spectra risk-score also added value in a model containing R-ISS, age at diagnosis and UAMS risk-score (APV = 0.109, $p = 7.4x10^{-5}$). In contrast, APV = 0.027 for UAMS in a model containing R-ISS, age at diagnosis and spectra risk-score. As our myeloma spectra are unsupervised, the same variable framework can be used to model any outcome or describe associations with clinical or demographic variables. For example, in multivariate Cox regression of time to treatment failure (TTF) 10 spectra were selected and predicted TTF with C_{adj} = 0.60 (0.59-0.66) and HR_{adj} = 3.10 (1.31-5.46) between high/low TTF risk-groups. In summary, the SPECTRA approach provides quantitative measures of transcriptome variation to deeply profile tumors. Application to MM outperformed UAMS and provides greater flexibility to describe tumors and model clinical outcomes.

PrgmNr 1362 - Integrated somatic and germline assessment of PROACTIVE genomic data

View session detail

Author Block: **A. Ghazani**[1], D. R. Koeller[2], A. Schwartz[2], D. K. Manning[3], H. Q. Rana[2], J. E. Garber[2]; [1]Brigham & Women's Hosp/Harvard Med. Sch., Boston, MA, [2]Dana Farber Cancer Inst, Boston, MA, [3]Brigham and Women's Hosp., Boston, MA

Disclosure Block: **A. Ghazani:** None.

Germline variants with low penetrance or in patients with non-syndromic phenotype present unique challenges in the interpretation of variant pathogenicity. In cancer, this challenge may be tempered by the presence of a second set of genomic information: the somatic genome. Here we present the results from genetic analysis of patients at Dana-Farber Cancer Institute and describe an integrated approach for the interpretation of somatic and germline findings. Germline results were obtained from DNA extracted from peripheral blood of patients enrolled in the PROACTIVE study, an institute-wide study that offers hereditary cancer testing of 133 genes for participants meeting disease-specific eligibility criteria. Somatic profiling of tumors was performed by OncoPanel at Brigham and Womenâ□□s Hospital to interrogate 447 cancer genes. To date, 913 patients have been enrolled in the PROACTIVE study; they are 43.9% male and 56.1% female aged between 6 months and 95 years old. Of those, 138 received reports that collectively contained deleterious germline variants in 45 genes; 24 genes (53%) are associated with high penetrance cancer syndromes and/or known clinical management guidelines. To date, 181 of patients have also had somatic sequencing obtained from the extracted DNA of the affected tumors. Reported somatic alterations include variants with well-established clinical utility (34), actionable gains (25) and losses (20), and structural rearrangements (25). Patients who were positive for germline deleterious variants in high penetrant genes but exhibited non-syndromic phenotypes or phenotypes discordant with the genotype were evaluated further using personal and family history of cancer as well as their somatic profiles in their component tumors when available. To date, we have identified ten probands and family members by this approach. Three germline *VHL* variants were found in those without a VHL syndromic phenotype; one germline *APC* variant was identified an individual without clinical suspicion of *APC* associated colon cancer or polyposis syndrome; one germline *MITF* variant was identified in an individual with vulvar mucosal melanoma. Concomitant evaluation of somatic and germline genome profiles, loss of function analysis, segregation, and detailed examination of phenotype and family history in these individuals suggested no evidence for penetrance in males or females in the germline *VHL* or *APC* families. The data also demonstrated new possible role of *MITF* beyond cutaneous melanoma. Our integrated approach of evaluation of both germline and somatic variants has led to the refined interpretation of germline variants and phenotypic implications.

PrgmNr 1363 - Germline HOXB13 variant contributes to risk of prostate cancer in men of African Ancestry

View session detail

Author Block: **R. Hughley**[1], B. F. Darst[2], C. Andrews[3], S. I. Berndt[4], X. Sheng[1], Z. Kote-Jarai[5], J. Lachance[6], C. W. Chiang[1], M. Jalloh[7], S. M. Gueye[7], A. O. Adebiyi[8], O. I. Aisuodionoe-Shadrach[9], B. Adusei[10], S. Mante[10], J. E. Mensah[11], A. A. Adjei[11], S. J. Chanock[4], S. Watya[12,13], R. Eeles[14,15], T. R. Rebbeck[16], D. V. Conti[1], C. A. Haiman[1], MADCaP Consortium, PRACTICAL Consortium, African Ancestry Prostate Cancer Consortium; [1]Ctr. for Genetic Epidemiology, Dept. of Preventive Med., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [2]Dept. of Preventive Med., Keck Sch. of Med., Univ. of Southern California,, Los Angeles, CA, [3]Harvard TH Chan Sch. of Publ. Hlth.and Div. of Population Sci., Dana Farber Cancer Inst., Boston, MA, [4]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, Bethesda, MD, [5]The Inst. of Cancer Res., London, United Kingdom, [6]Sch. of Biological Sci., Georgia Inst. of Technology, Atlanta, GA, [7]HÃ´pital GÃ©nÃ©ral Idrissa Pouye, Dakar, Senegal, [8]Univ. Coll. Hosp., Ibadan, Nigeria, [9]Univ. of Abuja and Univ. of Abuja Teaching Hosp., Abuja, Nigeria, [10]Dept. of Urology, 37 Military Hosp., Accra, Ghana, [11]Korle-Bu Teaching Hosp., Accra, Ghana, [12]Sch. of Publ. Hlth., Makerere Univ. Coll. of Hlth.Sci., Kampala, Uganda, [13]Uro Care, Kampala, Uganda, [14]The Inst. of Cancer Res., Sutton, Surrey, United Kingdom, [15]Royal Marsden NHS Fndn. Trust, London, United Kingdom, [16]Harvard TH Chan Sch. of Publ. Hlth., Boston, MA

Disclosure Block: **R. Hughley:** None.

Introduction: The non-synonymous germline HOXB13 G84E mutation is rare (0.1-1.4% frequency) in men of European ancestry and accounts for ~5% of hereditary prostate cancer. Recently, an African ancestry-specific germline deletion variant in HOXB13 (rs77179853), which removes the stop-codon (X285K), was found in Martinican men with early-onset prostate cancer (frequency=3.2%). Given the role of HOXB13 germline variation in prostate cancer, we investigated the association between the HOXB13 X285K variant and prostate cancer risk in a large sample of African ancestry men. Methods: We used freeze 8 of the TOPMed reference panel to impute the HOXB13 X285K variant in 22,361 men of African ancestry, which included 11,681 prostate cancer cases from the African Ancestry Prostate Cancer (AAPC) GWAS Consortium, the ELLIPSE/PRACTICAL OncoArray Consortium, the California/Uganda Prostate Cancer Study, the Ghana Prostate Study, and the Men of African Descent and Carcinoma of the Prostate (MADCaP) Consortium. Logistic regression models were used to assess the association of X285K with prostate cancer risk and aggressive disease adjusting for age, study, and the first ten principal components of ancestry. Results: The risk allele was only present in men of West African ancestry, with allele frequencies of 0.40% in controls from Ghana, 0.31% in controls from Nigeria, and 0% in controls and cases from Uganda and South Africa. Allele frequencies ranged from 0% to 0.26% in African ancestry controls from North America, the UK, and France, likely due to admixture. Limiting to populations where the variant was observed (10,476 cases and 9,687 controls), the HOXB13 X285K variant was significantly associated with a 2.4-fold increased risk of prostate cancer (95% CI=1.5-3.9, P=2.0x10-4; allele frequency in cases=0.35% and controls=0.14%). The variant was more common in 3,092 cases with high- or very-high risk disease (defined as Gleason 8-10, stage T3/T4, PSAâ¥20 nm/ml, metastatic disease, or died due to prostate cancer), with an allele frequency of 0.53% (OR=3.3, 95% CI=1.9-5.6, P=1.6x10-5, compared to controls). Conclusion: The HOXB13 X285K variant is associated with increased risk of prostate cancer in men of African ancestry. The variant is only found in men with West African ancestry and arose on an ancestral haplotype, although studies in other populations are needed to understand the distribution of the variant in Africa. Further analysis is needed to understand how the HOXB13 X285K impacts the HOXB13 protein and the function of this homeobox transcription factor in the prostate. Men carrying this variant may benefit from earlier PSA screening.

PrgmNr 1364 - Identification of mitochondrial DNA variants associated with risk of neuroblastoma

View session detail

Author Block: **X. Chang**[1], H. Hakonarson[2]; [1]CHOP, Philadelphia, PA, [2]Children S Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **X. Chang:** None.
Neuroblastoma is a childhood cancer that originates in the developing sympathetic nervous system. Emerging evidence indicate an important role of mitochondrial DNA (mtDNA) in the pathobiology of neuroblastoma. In this study, we aimed at identifying mtDNA variants associated with neuroblastoma risk. We applied a mitochondrial genome imputation pipeline to the SNP array data of two pediatric cohorts containing a total of 2,404 Caucasian children diagnosed with neuroblastoma and 9,310 ancestry-matched controls recruited at The Childrenâ□□s Hospital of Philadelphia. We further conducted case-control studies to explore potential associations of mtDNA variants with the susceptibility of neuroblastoma using logistic regression. Our results indicate that the mtDNA variant rs2853493 is significantly associated with neuroblastoma risk in the discovery cohort (odds ratio 0.62, P = 3.06 â□□ 10^{-9}). This result was further confirmed in the replication cohort (odds ratio 0.75, P = 2.43 â□□ 10^{-3}). Meta-analysis of the discovery and replication data identified rs2853493 (odds ratio 0.67, P = 9.36 â□□ 10^{-11}) and rs2853499 (odds ratio 0.71, P = 2.57 â□□ 10^{-8}) reaching the genome-wide significance level. Further, eQTL analysis shows that genotypes of the rs2853493 variant are associated with expression levels of MT-ND4 and MT-ND5 in multiple human tissues, and associated with MT-CYB gene expression in neuroblastoma cells. In this study we uncovered two mtDNA variants associated with neuroblastoma risk and provide information on candidate genomic regions and genetic markers of biological interest we propose to target in future studies.

PrgmNr 1367 - The genomic landscape of mosaic Y chromosome loss in the VA Million Veteran Program

View session detail

Author Block: **B. Gorman**[1], T. B. Bigdeli[2,3], G. Genovese[4], Y. Shi[1], S-G. Ji[5], C. Tcheandjieu[6,7], D. Klarin[8], A. Bick[9], S. M. Zekavat[10,4], P. Natarajan[8,4], V. Napolioni[11], R. P. Igo[12,13], T. L. Assimes[6,7], P. S. Tsao[7,6], N. S. Peachey[14,13], S. K. Iyengar[12,13], S. Pyarajan[1], Million Veteran Program; [1]VA Boston Hlth.care System, Boston, MA, [2]VA New York Harbor Hlth.care System, Brooklyn, NY, [3]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [4]Broad Inst., Cambridge, MA, [5]BridgeBio Pharma, Inc., Palo Alto, CA, [6]Stanford Univ. Sch. of Med., Stanford, CA, [7]VA Palo Alto Hlth.care System, Palo Alto, CA, [8]Massachusetts Gen. Hosp., Boston, MA, [9]Vanderbilt Univ. Med. Ctr., Nashville, TN, [10]Yale Sch. of Med., New Haven, CT, [11]Stanford Univ. Sch. of Med., Palo Alto, CA, [12]Case Western Reserve Univ., Cleveland, OH, [13]Louis Stokes Cleveland VA Med. Ctr., Cleveland, OH, [14]Cole Eye Inst., Cleveland Clinic, Cleveland, OH

Disclosure Block: **B. Gorman:** None.

Background: Mosaic loss of the Y chromosome (mLOY) in peripheral blood cells is implicated in aging and mortality-related disease processes in men (Forsberg, *Hum. Genet.* 2017). The Million Veteran Program (MVP) is a large, ancestrally diverse cohort, and a majority of participants are older men, making it ideal for studying mLOY.

Methods: Following recent advances (Thompson et al, *Nature* 2019), we called mLOY events in MVP by detecting shifts in allelic ratio in the pseudoautosomal region 1 (PAR1) using MoChA. We found detectable levels of mLOY in 26% of European-ancestry men, and 13% of African-ancestry and Hispanic-ancestry men.

Results: Genome-wide association scans (GWAS) identified 330 independent signals associated with mLOY in European-ancestry men (N=401,035), more than double the number of previously discovered signals. A GWAS in African-ancestry men (N=99,103) yielded 51 independent significant loci, including novel African-specific signals at *MPL*, *BUB1B*, *ETV6*, *SRCAP*, *TNRC6C*, and *XPC*. Finally, a GWAS in Hispanic/Latino men (N=44,046) yielded 18 signals. We refined these associations through local ancestry deconvolution (Tractor), trans-ancestry fine-mapping, and colocalization analyses. We additionally found 18 novel rare variant associations with mLOY, including germline mutations in *CHEK2* and *ATM* and somatic mutations in *DNMT3A*, *IDH2*, and *JAK2*, all genotyped on the MVP array. Leveraging the VAâ□□s electronic health record, we replicated previously reported associations with mLOY and additionally observed a striking negative association between metabolic phenotypes and mLOY, which was robust to adjustment for blood cell counts. We subsequently demonstrated this relationship to be causal (metabolic syndrome -> reduced mLOY) using summary-based Mendelian randomization experiments.

Conclusions: We conducted the first trans-ancestry analysis of mLOY and discovered numerous novel variant associations in European and African ancestries. As MVP tends to be older and with more comorbidities than other similarly sized cohorts, we found novel associations with metabolic traits, highlighting connections between the immune and neuroendocrine systems.

PrgmNr 1368 - Genome-wide gene expression analyses reveals ancestry-specific genetic architecture in Latino and African American children

View session detail

Author Block: **A. Mak**[1], L. Kachuri[2], D. Hu[1], C. Eng[1], S. Huntsman[1], J. Elhawary[1], N. Gupta[3], S. B. Gabriel[3], S. Xiao[4], H. Gui[4], K. Williams[4], J. RodrÃguezâ□□Santana[5], M. LeNoir[6], K. L. Keys[1], A. O. Oni-Orisan[1], S. S. Oh[1], M. A. Seibold[7], C. R. Gignoux[8], N. Zaitlen[9], E. G. Burchard[1], E. Ziv[1]; [1]Univ. of California, San Francisco, San Francisco, CA, [2]UCSF, San Francisco, [3]Broad Inst., Cambridge, MA, [4]Henry Ford Hlth.System, Detroit, MI, [5]Centro de NeumologÃa PediÃitrica, San Juan, PR, [6]Bay Area Pediatrics, Oakland, CA, [7]Natl. Jewish Hlth., Denver, CO, [8]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [9]Univ. of California, Los Angeles, Los Angeles, CA

Disclosure Block: **A. Mak:** None.

Background: Non-European populations are under-represented in both genome-wide association studies (GWAS) and expression quantitative trait loci (eQTLs) reference databases. The lack of adequate eQTL datasets limits fine mapping of GWAS results and the application of transcriptome-wide association studies (TWAS) in non-European ancestry populations. We leveraged whole genome and RNA sequencing data from 2,280 African American, Mexican American, and Puerto Rican children with and without asthma to investigate the relationship between genetic ancestry and heritability of gene expression. We quantified the prevalence of ancestry-specific eQTLs, developed gene expression models for TWAS, and demonstrated the gains in predictive power.

Results: Heritability (h^2) of gene expression in *cis* was highest in participants with the higher African (AFR) ancestry populations and lowest in participants with the higher Indigenous American ancestry (IAM). Participants with >50% AFR (AFR_{high}: h^2=0.17) had significantly higher h^2 compared to individuals with low: h^2=0.13, p=2.0Ã□10^{-147}). Among participants with >50% IAM, heritability was lower (h^2=0.12) compared to h^2=0.16, p=2.0Ã□10^{-147}). The results for higher heritability in AFR and lower for IAM were consistent when we used locus specific ancestry for heritability comparisons. We developed a framework to identify ancestry-specific eQTLs, accounting for linkage disequilibrium. We studied 9,635 heritable genes in the AFR_{high} individuals and found that over 25% had ancestry-specific eQTLs. We generated gene expression imputation models for 11,807 genes (mean cross-validation R^2=0.16) and compared these models with models based on GTEx and the Multi-Ethnic Study of Atherosclerosis (MESA) in a TWAS of 28 traits from the Population Architecture using Genomics and Epidemiology (PAGE) Consortium. The total number of genes examined in our models was 38% to 53% higher than GTEx and MESA, respectively. Applying our models to multi-ancestry GWAS results from PAGE identified 321 significantly associated genes (FDR-3) and in GTEx (in 83% of analyses, p=7.5Ã□10^{-3}).

Discussion: We found that *cis*-heritability of gene expression tracked with heterozygosity (highest in AFR and lowest in IAM). We also found that ancestry-specific eQTLs are common in a large fraction of genes, stressing the need for larger GWAS and RNA-seq sample size in AFR and IAM populations. Finally, we demonstrated the improved performance of ancestry-specific gene expression models for gene discovery in populations with mixed ancestry.

PrgmNr 1369 - Single-cell RNA-sequencing reveals pervasive but highly cell type-specific genetic ancestry effects on the response to viral infection

View session detail

Author Block: **H. Randolph**[1], J. K. Fiege[2], B. K. Thielen[3], M. S. Cobb[4], J. Barroso-Batista[4], R. A. Langlois[2], L. B. Barreiro[5]; [1]Genetics, Genomics, and Systems Biology, Univ. of Chicago, Chicago, IL, [2]Microbiol. and Immunology, Univ. of Minnesota, Minneapolis, MN, [3]Dept. of Pediatrics, Univ. of Minnesota, Minneapolis, MN, [4]Genetic Med., Univ. of Chicago, Chicago, IL, [5]Univ. of Montreal/CHU Ste-Justine, Montreal, QC, Canada
Disclosure Block: **H. Randolph:** None.
Humans vary in their susceptibility to infectious disease, partly due to variation in the immune response following infection. Genome-wide association studies and eQTL mapping studies in immune cells have shown that certain polymorphisms drive variation in the response to influenza, including expression patterns of key antiviral regulators. Influenza infection has also been shown to induce marked differences in response that are correlated with genetic ancestry in monocytes derived from African and European individuals. Yet, relatively little is known about the underlying genetic factors that contribute to heterogeneity in the influenza response across different immune cell types and subsets. Here, we used single-cell RNA-sequencing to quantify genetic contributions to this variation in peripheral blood mononuclear cells across 90 individuals, focusing specifically on the transcriptional response to influenza infection. We find that monocytes are the most responsive to infection, but that all cell types mount a conserved interferon response, which is stronger in individuals with increased European ancestry. By comparing expression patterns between European American (EA) and African American (AA) individuals, we show that genetic ancestry effects on expression are common, influencing 29% of genes, but highly cell type-specific, with over half detected in only one or two cell types. Further, we demonstrate that, on average, 53% of population-associated expression variation is explained by *cis*-expression quantitative trait loci. Finally, we provide evidence that genes associated with COVID-19 severity are strongly enriched for genes differentially expressed between African- and European-ancestry individuals, suggesting that immune response variation may compound some of the known health disparities that contribute to differences in COVID-19 susceptibility reported between AA and EA individuals. Our findings establish common *cis*-regulatory variantsâ□□including those that are differentiated by genetic ancestryâ□□as important determinants of the antiviral immune response.

PrgmNr 1370 - Chromatin accessibility differences between African and European Alzheimer Disease brains

View session detail

Author Block: **K. Celis**[1], A. J. Griswold[1], F. RAJABLI[2], P. Whitehead[1], K. L. Hamilton-Nelson[3], D. M. Dykxhoorn[1], K. Nuytemans[4], L. Wang[1], M. E. Flanagan[5], M. Mesulam[6], S. Weintraub[7], C. Geula[6], M. Gearing[8], G. Klein[9], D. Bennett[9], M. A. Pericak-Vance[10], J. Young[2], J. M. Vance[1]; [1]Univ. of Miami, Miami, FL, [2]Univ Miami, Miami, FL, [3]Univ. of Miami Miller Sch. of Med., Miami, FL, [4]John P. Hussman Inst. for Human Genomics, Miami, FL, [5]Northwesten Univ., Chicago, IL, [6]Northwestern Univ., Chicago, IL, [7]Northwestern Univ., Chigaco, IL, [8]Emory Univ., Atlanta, GA, [9]Rush Univ. Med. Ctr., Chicago, IL, [10]Miami, FL

Disclosure Block: **K. Celis:** None.

Background: Local ancestry (LA) surrounding the APOE4 allele is associated with an increased risk for Alzheimer Disease (AD) in European (EU) LA compared to African (AF) LA. We recently demonstrated that APOE4 has significantly higher expression in astrocytes from EU compared to AF APOE4 LA carriers, but the specific molecular mechanism leading to this difference is not known. We investigated whether evaluating chromatin accessibility in the LA region in prefrontal cortex from both AF and EU LA could lead us to identify potential regulatory factors affecting chromatin remodeling and expression of APOE. **Method:** We performed single nuclei Assays for Transposase Accessible Chromatin sequencing (snATAC-seq) in five AF LA and six EU LA individuals with AD, all homozygous APOE4/4 carriers. We processed the snATAC-seq data using Archr. **Result:** In total, 46,454 nuclei were sequenced with an average of 11,698 median fragments per nuclei. Overall, 729,589 chromatin accessible peaks were identified. We detected 24 clusters using UMAPhCluster resolution 1.3. Most of the clusters showed no differential accessibility when comparing ancestry groups, however four clusters had increased accessibility in AF and two had decreased accessibility in AF. Within the two clusters with decrease accessibility in AF, the astrocytic cluster 11 had the most differentially chromatin accessible peaks between AF and EU (325 vs 7895 peaks respectively). Interestingly, 1,045 differentially accessible peaks were detected in chromosome 19, from which 56 peaks fall in the LA region. The APOE local ancestry region showed reduced accessibility in AF when compared to EU. Additionally, around 44% of the differentially accessible sites in cluster 11 are in promoter regions of genes. Further, enrichment analysis of genes with altered chromatin accessibility at their promoter regions showed enrichment for Alzheimer disease pathways. **Conclusion:** This study is one of the first studies to compare chromatin accessibility in AF and EU ancestry brains. It also represents initial efforts to investigate potential underlying mechanisms that contribute to the differential expression we have previously reported in APOE4 between AF and EU LA brains. Our data suggest that decreased chromatin accessibility around APOE in astrocytes correlate with the observed decreased expression of APOE4 from AF LA groups.

PrgmNr 1371 - Protein prediction for trait mapping in diverse populations

View session detail

Author Block: **R. Schubert**[1], I. Gregga[1], E. Geoffroy[1], A. Mulford[1], A. W. Manichaikul[2], H. Im[3], H. E. Wheeler[4], NHLBI TOPMed Consortium; [1]Loyola Univ. Chicago, Chicago, IL, [2]Univ Virginia, Charlottesville, VA, [3]The Univ. of Chicago, Chicago, IL, [4]Loyola Univ Chicago, Chicago, IL
Disclosure Block: **R. Schubert:** None.
Genetically regulated gene expression has helped elucidate the biological mechanisms underlying complex traits and similar interrogation of the proteome is now possible. Here, we used the Trans-omics for Precision Medicine (TOPMed) Multi-omics pilot study, which comprises data from participants in the Multi-Ethnic Study of Atherosclerosis (MESA) cohort, to optimize genetic predictors of the plasma proteome for genetically regulated proteome association studies.
For 1305 proteins measured by a SOMAscan assay, we compared predictive models built via baseline elastic net regression to models integrating posterior inclusion probabilities estimated by fine-mapping SNPs prior to elastic net. In order to investigate the transferability of predictive models across ancestries, we built protein prediction models in five race/ethnic groups from MESA: African American (AFA, n = 183), Chinese (CHN, n = 71), European (EUR, n = 416), Hispanic/Latino (HIS, n = 301), and all populations combined (ALL, n=971).
We successfully built predictive models for 1187 unique proteins (R>0.1). As expected, fine-mapping produced more protein prediction models. Despite differences in sample size, EUR, HIS, and AFA training populations produced comparable numbers of predictive models. We used INTERVAL (n=3,301), a European ancestry study for out of sample estimation of model performance. For the proteins predicted by both the ALL and EUR training populations in INTERVAL, the ALL population predicted better than EUR with both the baseline (p=0.0012) and fine-mapped (p=0.0064) model building strategies. At current training population sample sizes, performance between baseline and fine-mapped protein prediction models was similar.
Using GWAS summary statistics from the Population Architecture using Genomics and Epidemiology (PAGE) study, which comprises â¼50,000 Hispanic/Latinos, African Americans, Asians, Native Hawaiians, and Native Americans, we applied S-PrediXcan to perform proteome association studies for 28 complex traits. The most protein-trait associations were discovered, colocalized, and replicated using proteome model training populations with similar ancestries to PAGE (i.e. predominantly African American and Hispanic). These 21 distinct associations provide more evidence that the SNPs at the locus are acting through protein abundance regulation to affect the associated phenotype and include: HP and Apo E associated with cholesterol traits and CRP, IL-1Ra, IL-6 sRa, and Apo E associated with C-reactive protein. More omics data in diverse populations are needed to better understand the mechanisms underlying complex traits in all populations.

PrgmNr 1372 - Multi-ancestry GWAS for venous thromboembolism identifies novel loci followed by experimental validation

View session detail

Author Block: **B. N. Wolford**[1], K-H. Wu[1], I. Surakka[2], Y. Zhao[3], X. Yu[3], C. Richter[3], K. C. Desch[2], M. Daly[4], C. J. Willer[1], J. Shavit[5], W. Zhou[6], Global Biobank Meta-analysis Initiative; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ MICHIGAN, Ann Arbor, MI, [3]Div. of Pediatrics Hematology/Oncology, Michigan Med., Ann Arbor, MI, [4]Harvard Med. Sch., [5]Div. of Pediatrics Hematology/Oncology, Ann Arbor, MI, [6]Massachusetts general Hosp., Broad Inst., Boston, MA

Disclosure Block: **B.N. Wolford:** None.

Deep vein thrombosis and pulmonary embolism, collectively referred to as venous thromboembolism (VTE), are characterized by the formation of thrombi in large veins that may embolize to the pulmonary circulation. VTE is a common cause of morbidity and mortality, and while driven by environmental and lifestyle factors, is also estimated to be 40% heritable. After phenotype harmonization, genome wide association study (GWAS) results across 9 international cohorts of the Global Biobank Meta-analysis Initiative (GBMI), with representation across 6 super populations (NFE, AMR, AFR, EAS, FIN, SAS; cases=27,987, controls=1,035,290), were combined using inverse-variance weighted meta-analysis. This multi-ancestry GWAS resulted in 38 genome-wide significant loci, 22 of which were potentially novel. For each autosomal locus we performed gene prioritization using 6 independent, yet converging, lines of evidenceâ□□presence of an eQTL in relevant GTEx tissues, pathogenic variant in ClinVar, nonsynonymous variant in 95% credible set (calculated with sum of single effects [SuSiE]), the nearest gene, in genes with p-value

PrgmNr 1375 - Quantifying the developmental trajectory of autism associated brain overgrowth using 3D cellular resolution imaging

View session detail

Author Block: **F. Kyere**[1], C. M. McCormick[2], O. Krupa[2], M. R. Glass[2], T. M. Farah[2], J. A. Jimenez[2], M. S. Itano[2], M. J. Zylka[2], G. Wu[2], J. Stein[3]; [1]NeuroSci., UNC Chapel Hill, Chapel Hill, NC, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]UNC Chapel Hill, Chapel Hill, NC
Disclosure Block: **F. Kyere:** None.
Brain development involves the organized differentiation of neural progenitors into neurons and glia, tightly orchestrated in both temporal and spatial domains. Alterations in embryonic brain development can manifest as altered postnatal brain structure and function, leading to neuropsychiatric illness. One such alteration is mutation in *Chromodomain helicase DNA binding protein 8 (ChD8)*; a gene that encodes an ATP-dependent chromatin remodeler. Heterozygous *Chd8* loss of function mutations result in ASD and macrocephaly in both humans and mouse models. However, how *Chd8* haploinsufficiency impacts brain structure across cortical development is unknown. Importantly, the increase in brain volume is localized in specific brain regions including the cortex, requiring a whole brain imaging approach to study the mutation effects. Here, we employ tissue clearing technology, light-sheet and confocal microscopy to acquire 3D cellular resolution images in intact whole brains derived from a mouse model with a single heterozygous point mutation in *ChD8*. Our goal is to quantify, using computational tools, neural progenitor and neuronal cell-types within annotated areas of the developing neocortex across critical time periods of neocortical neurogenesis in both WT and $Chd8^{+/-}$ mice. With this we will be able to determine the cell types that drive this ASD associated cortical hyper-expansion. At present, we have generated 3D cellular high resolution (0.75Âµmx0.75Âµmx4Âµm) images from a male $Chd8^{+/-}$ early postnatal time point (P4) labeled with upper layer neuronal marker (*Brn2*), lower layer neuronal marker (*Ctip2*), and a nuclear marker (YO-PRO). Using NuMorph, a suite of image analysis tools developed in our lab, we have registered the nuclei images to the Allen Developing Brain Atlas and quantified all the nuclei in the cortex. We detected about 13.8 million nuclei in the cortex which were YO-PRO positive. We also detected 2.94 million nuclei (21% of cortical nuclei) which were $Ctip2^{+}$/YO-PRO^{+} and additionally 2.94 million nuclei (21% of cortical nuclei) which were $Brn2^{+}$/YO-PRO^{+}. These results suggest that roughly about 42% of all cells in the cortex of P4 mouse brain may be excitatory neurons. Ongoing experiments will permit comparison to an age-matched WT littermate control. Moving forward, we aim to generate 9 biological replicates for each sex per genotype at this and other early developmental timepoints to elucidate the cellular basis and spatial localization of brain overgrowth in our mouse model, leading to a better understanding of how genetic variation can alter cortical development and alter risk for neurodevelopmental disorders.

PrgmNr 1376 - Genomic architecture of autism spectrum disorder from comprehensive whole-genome sequence annotation

View session detail

Author Block: **B. Trost**[1], B. Thiruvahindrapuram[1], A. J. S. Chan[1], W. Engchuan[1], E. J. Higginbotham[1], J. L. Howe[1], L. O. Loureiro[1], M. S. Reuter[1], D. Roshandel[1], J. Whitney[1], M. Zarrei[1], M. Bookman[2], M. Fiume[3], R. K. C. Yuen[1], J. Sebat[4], T. Frazier[5], D. Glazer[2], D. M. Hartley[5], S. W. Scherer[1]; [1]Hosp. for Sick Children, Toronto, ON, Canada, [2]Verily Life Sci., San Francisco, CA, [3]DNAstack, Toronto, ON, Canada, [4]UC San Diego, La Jolla, CA, [5]Autism Speaks, New York, NY

Disclosure Block: **B. Trost:** None.

Background: Autism spectrum disorder (ASD) is extremely heterogeneous in both clinical presentation and genetic architecture. Fully understanding ASD genetics requires whole-genome sequencing (WGS), which theoretically allows the genome-wide detection of all sizes and types of variants. Our objective is to leverage the power of WGS to assess the contribution to ASD of a wide range of variation, including common and rare, sequence-level and structural, and coding and non-coding. **Results:** Using sequence-level variants from the largest ASD sequencing dataset assembled to date (>18,000 trios plus additional cases and controls with exome or genome sequencing data), we used the previously-described TADA+ method to generate a list of ~140 ASD genes with FDR n=11,312, including 5,102 with ASD), the Simons Simplex Collection (*n*=9,205, including 2,419 with ASD), and the 1000 Genomes Project (*n*=2,504 population controls). Frequently detected genomic disorder syndromes included 15q11-q13 duplications, 16p11.2 deletions, and 1q21.1 duplications. ASD-risk genes commonly affected by SVs included *AUTS2*, *MBD5*, *NRXN1*, *SHANK1*, and *PTCHD1-AS*. We identified several ASD-relevant SVs that could only have been characterized with WGS, including an inversion disrupting *SCN2A* and a complex rearrangement impacting *KCNQ2*. Calculating polygenic risk scores (PRSs) using a recent ASD GWAS, polygenic risk was significantly over-transmitted from parents to ASD-affected children (p -7) but not to unaffected siblings. No difference in mean PRS was observed between ASD-affected children in multiplex families (where ASD is more likely to be caused by inherited rather than *de novo* variation) versus simplex, suggesting that rare inherited variation may have a more prominent role in multiplex ASD. ASD-affected individuals without a known clinically significant rare variant had significantly higher PRS. Despite these aggregate trends, PRS was not predictive at the level of individuals or families. Transmission tests involving non-coding genomic features suggested a role for the disruption of boundaries of topologically associating domains containing ASD-risk genes. **Conclusion:** By analyzing many types of genetic variants (also including tandem repeat expansions from a recently-published study), we present what we believe to be the most comprehensive analysis of genetic variation in ASD thus far, providing new insights into the genetic heterogeneity of ASD.

PrgmNr 1377 - Genotype-phenotype analyses in 10,000 individuals with psychosis highlights complex role of ultra-rare coding variants and common polygenic risk in presentation and etiology

View session detail

Author Block: **T. Singh**[1], A. Palotie[2], B. M. Neale[2], M. Daly[3], on behalf of the SUPER psychosis study; [1]Massachussetts Gen. Hosp., Boston, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Harvard Med. Sch.
Disclosure Block: **T. Singh:** None.
Psychiatric disorders, particularly in psychotic disorders like schizophrenia, have seen significant progress in loci discovery, with common regulatory variants and ultra-rare coding variants in individual genes identified at genome-wide significance. Despite this success, the degree to which clinical presentation is influenced by different types of genetic risk has not been fully explored, with such efforts hampered by limited data in current studies and exacerbated by the imprecision of diagnostic criteria across disorders. Here, we present the SUPER psychosis study, a comprehensive nation-wide effort in Finland to recruit 10,417 individuals diagnosed with schizophrenia, bipolar disorder or major depressive disorder with psychotic features. All individuals underwent standardized psychiatric interviews and questionnaires with accessible electronic health record data from a national registry. Combined with imputed genotype and whole-exome sequence data, we explore how genetic risk from common and rare variants are associated with psychiatric symptomatology, severity and progression, cognition, and additional comorbidities. We replicated a significant excess of rare protein-truncating variants (PTVs) in schizophrenia-associated genes (P = 2.1 x 10-9) in psychosis cases compared to 11,367 matched controls, with these PTVs observed in 90 (~0.9% of) cases. Notably, this rare variant signal was observed not exclusively in diagnoses of schizophrenia (P = 7.5 x 10-10), but also across the broader psychosis spectrum that includes bipolar disorder (P = 6 x 10-5) and unspecified/other psychoses (P = 0.00275). We present additional associations between genetic risk, psychiatric symptomatology, and additional comorbidities, highlighting the dramatic concentration of rare variant carriers in individuals with additional cognitive impairment and learning difficulties compared to those without (2.5% compared to 0.86% of cases). We explore the phenotypic range of individuals carrying PTVs in the same genes (9 SETD1A, 7 AKAP11, and 6 SRRM2), and find that certain genes resemble more typical psychiatric presentation while all six SRRM2 carriers had additional intellectual disability. Combined, we show the immense value of standardized and comprehensive phenotyping when building new collections for advancing gene discovery, especially when studying highly heterogeneous syndromes and disorders.

PrgmNr 1378 - Genome-wide association study of over 40,000 cases of suicide attempts yields eight genome-wide significant loci

View session detail

Author Block: **A. Ashley-Koch**[1], N. Kimbrel[1], X. J. Qin[1], J. H. Lindquist[2], M. E. Garrett[1], R. K. Madduri[3], J. E. Huffman[4], H. Coon[5], A. R. Docherty[5], N. Mullins[6], D. M. Ruderfer[7], B. H. McMahon[8], E. R. Hauser[9], M. A. Hauser[1], J. C. Beckham[2], International Suicide Genetics Consortium, Million Veterans Program Suicide Exemplar Workgroup; [1]Duke Univ., Durham, NC, [2]Durham VA, Durham, NC, [3]Argonne Natl. Lab, Lemont, IL, [4]VA Boston Hlth.care System, Jamaica Plain, MA, [5]Univ. of Utah, Salt Lake City, UT, [6]Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]Vanderbilt, Nashville, TN, [8]Los Alamos Natl. Lab, Washington, DC, [9]Duke Univ Med Ctr, Durham, NC

Disclosure Block: **A. Ashley-Koch:** None.

Background Suicide accounts for >800,000 deaths per year worldwide and non-fatal suicide attempts occur over 20 times more frequently. Suicide attempt (SA) is heritable and only partially overlaps with the genetic etiology of related psychiatric disorders. The Million Veteran Program (MVP) and the International Suicide Genetics Consortium (ISGC) have conducted the two, independent largest genome-wide association studies (GWAS) of SA to date, identifying several genome-wide significant loci. Due to the polygenic etiology of SA, further increases in sample size should yield additional genetic loci. Here, we present a GWAS meta-analysis of SA between the MVP and ISGC cohorts, including >40,000 SA cases. **Methods** The MVP cohort included 14,089 veterans who made a SA and 395,359 controls. SA was defined from ICD codes, mental health surveys and data from the Suicide Prevention Application Network database. The ISGC cohort included 29,782 individuals who made a lifetime SA and 519,961 controls from 18 studies worldwide. SA was defined by psychiatric interviews for 10 studies, and self-report, hospital records, or coronersâ reports in remaining studies. GWAS of SA were conducted in each cohort and the genetic correlation between them was calculated using LD score regression (LDSC). A trans-ancestry fixed-effects meta-analysis was conducted between the cohorts, including 43,871 SA cases and 915,320 controls. Amongst SA cases, 81% were of European ancestry, with 11% African American, 5% Asian and 3% Latinx ancestry. LDSC was used to calculate the SNP-heritability of SA. Enrichment of SA-association signal in biological pathways and tissues was explored using FUMA. **Results** The genetic correlation of SA between the ISGC and MVP cohorts was 0.86 (se 0.09) and not significantly different from 1. Trans-ancestry meta-analysis found 8 loci reaching genome-wide significance for SA (PSLC6A9, NLGN1, ESR1, DRD2 and *FURIN* genes. The SNP-heritability of SA was 3.8% (se 0.003) on the liability scale. Genetic associations with SA were significantly enriched in genes expressed in brain tissues from the Genotype-Tissue Expression (GTEx) project. **Discussion** This meta-analysis is the largest GWAS of SA to date and implicates novel risk loci and biologically relevant pathways and tissues. Ongoing work includes ancestry-specific GWAS meta-analyses, integration of GWAS results with â omics data from brain tissues, and drug target enrichment analyses. These analyses will provide further novel insights into SA etiology and treatment.

PrgmNr 1379 - Chromatin accessibility mapping reveals compound genetic effects at a schizophrenia GWAS risk locus impairing neurodevelopment and synaptic function in human neurons

View session detail

Author Block: **S. Zhang**[1], H. Zhang[1], M. Forrest[2], Y. Zhou[3], V. Bagchi[2], A. Kozlova[4], M. Dos Santos[2], N. H. Piguel[2], L. Dionisio[2], A. R. Sanders[1], Z. P. Pang[5], X. He[6], P. Penzes[2], J. Duan[7]; [1]North Shore Univ Hlth.System, Evanston, IL, [2]Northwestern Univ., Chicago, IL, [3]Univ. of Chicago, [4]The Res. Inst., NorthShore Univ. Hlth.System, Evanston, IL, [5]Robert Wood Johnson Med. Sch., Rutgers Univ., New Brunswick, NJ, [6]Univ Chicago, Chicago, IL, [7]Northshore Univ Hlth.system/Univ of Chicago, Evanston, IL
Disclosure Block: **S. Zhang:** None.
Genome-wide association studies (GWAS) of schizophrenia (SZ) and other neuropsychiatric disorders have identified hundreds of risk single nucleotide polymorphisms (SNPs). However, functional interpretation of these GWAS risk SNPs has been challenging. Recently, we described an approach to elucidate functional SNPs by their allele-specific open chromatin (ASoC) signatures (Zhang *et al.*, *Science* 2020). Here, using ASoC mapping and CRISPR/Cas9 genome editing in *NEUROG2* (*NGN2*)-induced excitatory neurons (NGN2-iNs), we aimed at making causal links between GWAS risk SNPs and SZ-relevant disease phenotypes. We initially identified thousands of ASoC SNPs in NGN2-iNs of 20 iPSC lines and found 31 credible functional SNP in 26 SZ risk loci (out of 108) that displayed ASoC. For the four SZ-associated ASoC SNPs (near *VPS45*, *BCL11B*, *PRBML1*/*GNL3*, and *BAG5*) that also replicated our previous ASoC results (Zhang *et al.*, *Science* 2020) in neural progenitor cells (NPCs)-derived glutamatergic neurons, we performed multiplex CRISPRi/single-cell RNA-seq (CROP-seq) in NGN2-iNs and found three SNP sites exhibited *cis*-regulatory effects on their flanking gene targets. Since the *VPS45* locus showed the strongest ASoC in ATAC-seq and transcriptional repression of *cis*-targets in CROP-seq, we further investigated the function of this ASoC SNP (rs2027349) by CRISPR/Cas9 editing. In NGN2-iNs derived from isogenic iPSC lines carrying the two different alleles of rs2027349, we found that risk allele A was associated with an increased expression of the bi-directionally transcribed genes *VPS45* and *AC244033.2* (the latter being a lncRNA), as well as a distal gene *C1orf54*. Notably, NGN2-iNs carrying risk allele A showed transcriptomic changes correlated with gene expression changes in post-mortem brains of SZ, bipolar disorder, and autism spectrum disorder. Phenotypically, NGN2-iNs with A allele exhibited increases in dendritic complexity, synaptic puncta density, neural firing rate, and the frequency of calcium transients. Moreover, RNA knockdown of each *cis*-regulated gene in risk allele-carrying neurons partially rescued the risk allele A-associated phenotypic changes, suggesting a phenotypic contribution from all three genes. The influence of rs2027349 was further corroborated by transcriptomic modelling, brain chromatin interaction (Hi-C) data, and the assay of local chromatin accessibility in isogenic NGN2-iNs. Our study thus reveals a complex gene regulation process at a single SZ GWAS locus that impairs neurodevelopment and synaptic function, providing a mechanistic link between a noncoding SZ risk variant and disease-related phenotypes.

PrgmNr 1380 - Lineage-specific analysis of epigenome and transcriptome changes in postmortem brains from Schizophrenia and Bipolar Disorder

View session detail

Author Block: **J. Bendl**[1], K. Girdhar[2], G. E. Hoffman[3], J. Fullard[1], P. Roussos[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York City, NY, [2]Mount Sinai, New York, NY, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY

Disclosure Block: **J. Bendl:** None.

Background: Regions of open chromatin house regulatory elements required to mediate cell-type and tissue-specific gene expression. Studies of human brain have shown that dysregulation of these regulatory mechanisms is associated with neuropsychiatric diseases. Here, we present the largest cell type and brain region-specific study of differential chromatin accessibility in SCZ and BD brains.

Methods: Using frozen postmortem tissue from 469 cases with SCZ, BD, and controls, we performed ATAC-seq to profile chromatin accessibility in neuronal and glial populations of cells, isolated by FACS from two different brain regions, i.e. dorsolateral prefrontal cortex and anterior cingulate cortex. We further performed phenotype-aware deconvolution of neuronal samples to glutamatergic and GABAergic neurons as well as glial samples to oligodendrocytes, astrocytes, and microglia. Then, we characterized epigenetics changes associated with SCZ and BD phenotypes. The availability of RNA-seq and whole-genome data for this cohort allowed us to measure the overlap between differentially regulated transcriptome and epigenome signatures and related pathways. Finally, we determined enhancer-promoter interactions, by utilizing additional omics in the brain tissue (Hi-C and H3K27ac ChIP-seq) and jointly analyzing with ATAC-seq based on the â□□activity-by-contactâ□□ (ABC) approach.

Results: We observed widespread differences in chromatin accessibility associated with SCZ and BD phenotypes, mostly in neuronal samples. While substantial changes between cases and controls can be seen in both assayed brain regions, dorsolateral prefrontal cortex is, in general, more affected than anterior cingulate cortex. The differentially accessible regions of open chromatin were enriched in disease-related pathways and regulated by distinct transcription factors. By integrating these results with SCZ and BD GWAS, we were able to shed light on the mechanisms of several risk loci.

Conclusions: This dataset provides a unique insight into molecular mechanisms underlying brain region and cell type-specific vulnerability to both SCZ and BD. Supported by NARSAD 27209, R01MH110921 and R01AG057440.

Poster Talks

PrgmNr 1407 - Cancer Risk C, a functional genomics test, is a sensitive, specific, and rapid diagnostic for Lynch syndrome

View session detail

Author Block: **I. Alim**[1], J. Loke[2], S. Yam[1], S. D. Klugman[3], A. S. Templeton[4], P. A. Newcomb[4], N. M. Lindor[5], R. K. Pai[6], M. A. Jenkins[7], S. Gallinger[8], H. Ostrer[9]; [1]Morgan And Mendel Genomics, Bronx, NY, [2]1300 Morris park Ave Ullmann 817, Bronx, NY, [3]Montefiore Medical Center, Bronx, NY, [4]Fred Hutch, Seattle, WA, [5]Mayo Clinic, Scottsdale, AZ, [6]Mayo Clinic, Phoenix, AZ, [7]Univ Melbourne, Carlton, Australia, [8]Univ of Toronto, Toronto, ON, Canada, [9]Albert Einstein College of Medicine, Bronx, NY
Disclosure Block: **I. Alim:** None.
Heritable mutations in the DNA mismatch repair (MMR) pathway cause Lynch syndrome (LS), a condition that significantly increases risk of colorectal and other cancers. LS mutations are commonly found in the *MLH1*, *MSH2*, *MSH6*, *PMS2* and *EPCAM* genes. Diagnosis of LS is reliant on gene panel sequencing. However, at least half have a variant of uncertain significance (VUS) that cannot be classified for pathogenicity or have no result that informs their diagnosis. Many LS patients are identified only after microsatellite instability testing of tumors is found to be high (MSI-H). We developed a diagnostic test, Cancer Risk C (CR-C), using flow-cytometry based functional variant assays (FVAs) to aid in the diagnosis of LS. Using patient-derived lymphoblastoid cell lines (LCLs) from the Colon Cancer Family Registry (CCFR) an initial cohort was established with either known pathogenic (n=20) or benign variants (n=20). FVAs were tested to identify the effects of pathogenic variants in MMR genes on the nuclear translocation of the MLH1 and MSH2 proteins and the nuclear phosphorylation of the ATM and ATR proteins in response to treatment with the alkylating agent, methylnitronitrosoguanidine. To differentiate pathogenic and benign variants, a risk classification score was developed based on a combination of MLH1, MSH2 and ATR assays. CR-C was 98% sensitive and 95% specific and could be completed within 48 hours. A second cohort of CCFR patient-derived cells with either pathogenic (n=40) or benign variants (n=40) was tested with CR-C and observed to have similar sensitivity and specificity, thus creating a new diagnostic for LS. Amongst those with VUS and MSI-H tumors (n=60), 72% were identified as having LS. This finding contrasted to the previously observed rate of 16% MMR pathogenic gene variants amongst patients with MSI-H tumors. Edited cells with pathogenic variants in MMR genes were rescued from pathogenic to benign when transfected with an expression plasmid containing wild-type cDNA and tested with CR-C. Direct comparison of matched whole blood samples and LCLs yielded comparable results with CR-C (r^2>0.9). Compared to cancer gene panel sequencing, CR-C is more accurate and rapid for diagnosing LS and can be performed on whole blood. When combined with gene rescue experiments in edited cells or LCLs, CR-C can be used to classify VUS as pathogenic or likely pathogenic.

PrgmNr 1408 - Does it make sense to report single pathogenic alleles in cancer susceptibility genes for childhood cancer patients? Results from the BASIC3 and Texas KidsCanSeq studies

View session detail

Author Block: **S. E. Plon**[1], L. R. Desrosiers[1], S. Scollon[2], A. K. Petersen[3], H. DAI[4], J. Reuther[5], G. Miles[6], D. M. Muzny[4], A. Roy[5], S. Kulkarni[7], R. A. Gibbs[4], A. L. McGuire[8], G. E. Tomlinson[9], J. Bernini[1], J. B. Gill[10], T. Griffin[11], K. Vallance[12]; [1]Department of Pediatrics, Baylor College Medicine, Houston, TX, [2]Baylor College of Medicine, Houston, TX, [3]Randall's Children Hospital and Legacy Health, Portland, OR, [4]Baylor College Medicine, Houston, TX, [5]Pathology, Baylor College Medicine, Houston, TX, [6]Molecular and Human Genetics, Baylor College Medicine, Houston, TX, [7]Baylor Genetics, Houston, TX, [8]Baylor Col Med, Houston, TX, [9]UT Health San Antonio, San Antonio, TX, [10]UT MD Anderson Cancer Center, Houston, TX, [11]Children's Hospital of San Antonio, San Antonio, TX, [12]Cook Children's Medical Center, Fort Worth, TX

Disclosure Block: **S.E. Plon:** Non-remunerative positions or influence such as officer, board member, trustee, or public spokesperson; Baylor Genetics.

Background: Multiple studies report that 8-13% of pediatric cancer patients have pathogenic or likely pathogenic variants (PV/LPV) in cancer susceptibility genes, although standards vary for reporting single PV/LPVs in genes associated with autosomal recessive (AR) syndromes. We report here results from two studies of pediatric cancer patients for the frequency of single AR variant results and relationship to the patientâ□□s presentation. **Methods:** The BASIC3 study included clinical germline and tumor exome sequencing in an ethnically diverse cohort of sequentially diagnosed children with CNS and non-CNS solid tumors. The Texas KidsCanSeq study included germline exome and panel sequencing for pediatric solid tumor patients across six sites in Texas. In both studies, other medical problems noted in the medical record were provided to the reporting laboratory. All result categories on the clinical exome and panel (when available) were returned to parents. **Results:** Among 278 enrolled patients in BASIC3 there was one patient with two PVs (a homozygous variant in *TJP2*) and 18 (6.5%) patients had a single PV/LPV in a wide spectrum of AR cancer genes. Only one patient had a tumor type (Wilms tumor) previously associated with the gene (*DIS3L2*). Patent foramen ovale in a patient with *FANCA* allele was the only syndromic feature. In the first 366 patients in KidsCanSeq there were no patients with biallelic PV/LPVs, 15 patients (4.1%) with one PV/LPV in AR genes of which 2 had the associated tumor type (*DIS3L2* with Wilms tumor and *SLC26A4* with thyroid cancer). Not included here are patients with PV/LPV in genes with both AR and dominant cancer phenotypes e.g., *ATM.* Germline results disclosure with parents discussed (1) their oncologist might request further targeted testing for a 2nd variant if there is concern that the patient has this rare disorder, (2) the patient (and often parent) are carriers of this rare disorder, (3) the lack of evidence relating the variant to the childâ□□s cancer diagnosis and (4) lack of any additional surveillance recommendations.**Conclusions:** There is increasing use of exome, genome or large hereditary cancer panels for genomic evaluation of cancer patients. Recessive diagnoses are very uncommon but about 5% of pediatric cancer patients have a single PV/LPV in rare AR cancer genes without other findings. These test results require substantial efforts for disclosure and family understanding. We suggest that limiting reporting to biallelic patients or those with features of the disorder including the specific cancer type would streamline the process at the cost of reproductive information. BASIC3 and Texas KidsCanSeq were supported by 1U01HG006485.

PrgmNr 1409 - Breast cancer polygenic risk scores and rare variants in Latinas

View session detail

Author Block: **J. L. Nierenberg**[1], A. Adamson[2], Y. C. Ding[2], Y. Shieh[1], D. Hu[1], S. Huntsman[1], E. M. John[3], G. Torres-Mejia[4], C. A. Haiman[5], L. H. Kushi[6], C. N. Ricker[7], L. Steele[2], R. Lee[1], J. N. Weitzel[8], L. Fejerman[9], S. L. Neuhausen[2], E. Ziv[1]; [1]University of California San Francisco, San Francisco, CA, [2]Beckman Research Institute of City of Hope, Duarte, CA, [3]Stanford University School of Medicine, Stanford, CA, [4]Instituto Nacional de Salud Pul□blica, Cuernavaca, Mexico, [5]University of Southern California, Los Angeles, CA, [6]Kaiser Permanente Northern California, Oakland, CA, [7]USC Norris Comprehensive Cancer Center, Los Angeles, CA, [8]Sierra Madre, CA, [9]University of California Davis, Davis, CA

Disclosure Block: **J.L. Nierenberg:** None.

Introduction: Polygenic risk scores (PRS), assembled from common single nucleotide polymorphisms (SNPs), can be used to predict breast cancer risk. Individually, pathogenic variants (PVs) in high and intermediate penetrance breast cancer susceptibility genes are rare but have large effects on disease risk. Few studies have examined PRS and PVs in susceptibility genes together, and most previous breast cancer genetics studies have been conducted in European ancestry populations. Here, we report findings on the combined effects of PRS and PVs in Latinas.

Methods: We conducted a pooled case-control analysis of breast cancer in Latinas from the San Francisco Bay Area, Los Angeles, and Mexico (1,776 cases and 1,589 controls). Case ascertainment included 432 participants from high-risk studies (age below 50, family history, or bilateral breast cancer) and 1,344 from general population studies. We assembled a 180-SNP PRS from known breast cancer SNPs (P-8). We determined presence of a rare PV in 9 known breast cancer risk genes (*ATM*, *BARD1*, *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *PTEN*, *RAD51C*, and *TP53*). We used multivariable logistic regression and area under the receiver operating characteristic curve (AUC) to examine the relationships between PRS, PV status, and breast cancer risk. All analyses were adjusted for age, study, and ancestry. Secondary analyses were stratified by age, family history, or indigenous ancestry above or below the median.

Results: Higher PRS was associated with higher risk of breast cancer, with an odds ratio of 1.6 (95% confidence interval [CI]: 1.5-1.7) per PRS standard deviation and an AUC of 0.61 (95% CI: 0.60-0.63). PVs were found in 125 case and 22 control participants. Having a PV was associated with a 5.9-fold (95% CI: 3.8-9.6) increased odds of breast cancer. The AUC for breast cancer improved to 0.64 (95% CI: 0.62-0.66) when PVs were added to the PRS model. Results were similar among 255 cases and 129 controls with family history of breast cancer and among those with indigenous ancestry above and below the median. Among 599 cases and 387 controls under age 50 years, the AUC for the PRS-rare variant model increased to 0.69 (95% CI: 0.65-0.72). Among all cases, those with a PRS below the median had a 1.9-fold (95% CI: 1.4-2.7) increased odds of having a PV.

Conclusion: We found that adding PV status to a PRS model improves prediction of breast cancer, indicating that there may be clinical utility in examining PRS and PVs together, especially among young women. The lower PRS among carriers of a PV is likely an effect of case ascertainment and could be useful in epidemiological design of new gene discovery efforts.

PrgmNr 1410 - Germline genetically predicted mammographic density is associated with breast tumor transcriptomic and proteomic features, CD8+ T cell infiltration, somatic mutational signatures, and tumor mutational burden in The Cancer Genome Atlas

View session detail

Author Block: **A. Francis**[1,2], T. Robinson[1,2], B. Lloyd-Lewis[3], Y. Liu[4], C. Peng[4,5], R. Tamimi[6], P. Kraft[4], P. Brennan[7], T. Gaunt[1,2], R. M. Martin[1,2,8], C. Relton[1,2], A. Gabriel[9], G. Gierach[10], J. McKay[9], S. Kar[1,2]; [1]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom, [2]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom, [3]School of Cellular and Molecular Medicine, University of Bristol, Bristol, United Kingdom, [4]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, [5]Channing Division of Network Medicine, Brigham and Womenâ□□s Hospital, Boston, MA, [6]Division of Epidemiology, Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, [7]Genetic Epidemiology Group, International Agency for Research on Cancer, Lyon, France, [8]University Hospitals Bristol and Weston NHS Foundation Trust, NIHR Bristol Biomedical Research Centre, University of Bristol, Bristol, United Kingdom, [9]Genetic Cancer Susceptibility Group, International Agency for Research on Cancer, Lyon, France, [10]Integrative Tumor Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD

Disclosure Block: **A. Francis:** None.

High mammographic density (MD) is one of the strongest risk factors for breast cancer, the most common cancer in women globally, but the molecular underpinnings of this association are poorly understood. Tumor cohorts with multi-omic phenotyping currently do not have linked MD data at scale, making it challenging to define the genomic landscape of breast tumors arising on a background of high MD. We took a new approach to address this challenge by integrating MD genome-wide association study (GWAS) data with The Cancer Genome Atlas (TCGA) breast tumor data. We built a polygenic score for MD using effect size estimates and allele information for the lead variants at 17 independent loci (P -8) from a GWAS of percent MD in 24,192 women of European ancestry. We performed sample and genotype quality control and imputation into the 1000 Genomes reference panel on germline genetic data from 721 female breast cancer cases of genetically inferred European ancestry from TCGA. We assigned each woman in this TCGA cohort, which lacks measured MD, her genetically predicted MD based on the MD polygenic score. We evaluated the association between germline genetically predicted MD and breast tumor genome-wide transcriptomic, proteomic, genomic, epigenomic, and immune traits in TCGA using linear (default) and quasi-Poisson (for overdispersed count data) regression models adjusted for age and stage at diagnosis, estrogen receptor status, and 10 genetic principal components, reporting associations at FDR DLL1, an estrogen-dependent Notch ligand previously implicated in breast oncogenesis ($P = 2 \times 10^{-6}$; top gene of 20,530 genes profiled by RNA-Seq). High genetically predicted MD was associated with increased breast tumor CD8+ T cell infiltration (P = 0.004) in our evaluation of 22 tumor immune infiltrates profiled by the CIBERSORT algorithm. Genetically predicted MD had positive and inverse associations with breast tumor MAPK ($P = 3 \times 10^{-4}$) and XRCC1 ($P = 4 \times 10^{-4}$) protein levels, respectively, of the 281 tumor proteins profiled by reverse-phase protein array, suggesting roles for mitogen-activated protein kinase signaling and DNA repair. High genetically predicted MD was associated with the mitotic clock-like, aging-related single base substitution mutational signatures 1 (P = 0.001) and 5 (P = 0.002) and with high tumor mutational load (P = 0.006 for association with non-silent mutations/Mb). Thus, we combined germline and somatic data to identify breast tumor molecular features associated with genetically predicted MD, with potential implications for breast cancer development and progression.

PrgmNr 1411 - Saturation Genome Editing of *PALB2* Reveals Functionally Abnormal Missense Variants

View session detail

Author Block: **I. Hill**[1,2], A. Patoski[1,2], K. de Leon[1,2], A. Mehrotra[1,2], N. Smith[1,2], J. Shendure[1,2,3], L. Starita[1,2]; [1]Department of Genome Sciences, University of Washington, Seattle, WA, [2]Brotman Baty Institute for Precision Medicine, Seattle, WA, [3]Howard Hughes Medical Institute, Seattle, WA
Disclosure Block: **I. Hill:** None.
Pathogenic variants in the *PALB2* gene are associated with an increased lifetime risk of developing breast, pancreatic, and ovarian cancer. However, the clinical utility of genetic testing for informing *PALB2*-associated cancer susceptibility is limited by a lack of evidence for interpreting individual variants. In fact, of the 3125 *PALB2* single nucleotide variants (SNVs) in ClinVar, 1850 (60%) are currently annotated as variants of uncertain significance (VUS). The situation is worse for missense variants, of which 1849 of 1874 (99%) are VUS. Multiplexed assays of variant effects can provide valuable evidence for reinterpreting VUS. For example, functional data from Saturation Genome Editing of *BRCA1* can be used as strong evidence that individual variants are pathogenic (PS3) or benign (BS3). Like *BRCA1*, *PALB2* - **P**artner and **L**ocalizer of **B**RCA**2** - which functions in the same double strand break repair pathway as *BRCA1*, is compatible with Saturation Genome Editing.
We performed Saturation Genome Editing on *PALB2* to generate functional data for all possible SNVs across the protein coding region and intron-exon junctions. Thus far, we have data for 2602 variants, spanning the coiled coil domain which is required for BRCA1 binding and the WD40 domain which is required for binding to BRCA2. Functional scores indicate that 176 of 1408 (12.5%) missense variants are indistinguishable from nonsense variants and are likely â□□functionally abnormalâ□□, loss-of-function variants. Our results are consistent with what is known about PALB2 function, with missense variants scoring as functionally abnormal falling in the coiled-coil (10.5%) or WD40 domain (15.6%), and 0% outside of those domains thus far (n.b. there is more missing data outside of the folded functional domains). Our results are also perfectly concordant for 26 variants with published orthogonal functional data. Nonetheless, it should be recognized that the relative lack of classified pathogenic or benign missense variants within *PALB2* presents a unique and challenging problem in validating these data for use in clinical variant interpretation. New strategies may be necessary before these data can be incorporated into variant interpretation frameworks. Finally, recent case-control studies suggest that *PALB2* missense variants do not appear to increase breast cancer risk, however, our data suggests grouping all missense variants for these studies, of which 87.5% are â□□functionally normalâ□□, may have obscured any signal.

PrgmNr 1412 - Identify non-mutational TP53 loss of function in human cancers

View session detail

Author Block: **L. Wang**[1], Q. Li[2], Y. Zhang[2], L. Ma[2], Z. Zhang[2]; [1]Mayo Clinic, Rochester, MN, [2]Beijing Inst. of Genomics, Beijing, China
Disclosure Block: **L. Wang:** None.
Background: A gene with an intact DNA sequence can compromise its function by epigenomic, transcriptomic, and proteomic level dysregulations. Such non-mutational inactivations are prevalent in cancers but, due to the heterogenetic causes, cannot be detected by the DNA-sequencing, immunohistochemical staining, or any other single assay alone. Therefore, they would become a significant impediment for molecular diagnostics, clinical management, and treatment selection for cancer patients. **Approach**: We hypothesized that when a transcription factor (TF) is functionally impaired, either due to genetic or non-genetic causes, the expression of downstream target genes would be significantly altered and that such expression alteration, in turn, can be used to predict the functional status of the TF. Here we used p53 as an example; we first define the p53 target genes through a comprehensive literature review and meta-analysis. Then, we build an SVM model using the composite expression scores of these target genes as features and using the â□□normal tissuesâ□□ (assuming p53â□□s tumor suppressor function is normal in this group) and â□□*TP53* truncating tumor samplesâ□□ (assuming p53â□□s tumor suppressor function is lost in this group) as training datasets. **Results**: Using 5-fold cross-validation, we demonstrated the superior performance of our SVM model (average AUC = 0.995, F1-score = 0.989, recall = 0.992) in TCGA LUNG and BRCA cohorts. When applying our SVM model to *TP53*^{WT} tumor samples, we found 87% of BRCA and 94% of LUNG samples were predicted to be LoF (termed as *TP53*^{WT}-LoF). These *TP53*^{WT}-LoF patients exhibited distinct genomic and clinical characteristics from the other *TP53*^{WT} patients. Specifically, *TP53*^{WT}-LoF patients have significantly higher tumor mutation burden, the fraction of genome with copy number variations, aneuploidy score, and hypoxia score, consistent with p53â□□s function as a central regulator of DNA damage repair and cellular stress response. In addition, *TP53*^{WT}-LoF patients with lung cancer have significantly shortened overall survival compared to those real *TP53*^{WT} patients. Further analyses revealed that *MDM2*/*MDM4* amplifications are significantly enriched in *TP53*^{WT}-LoF patients, partially and mechanistically explained the p53 loss-of-function.

PrgmNr 1413 - Multiple germline events lead to cancer development in patients with Li-Fraumeni syndrome

View session detail

Author Block: **V. Subasri**[1,2,3], N. Light[1,2], N. Kanwar[1], J. R. Hansford[4], K. E. Nichols[5], K. C. de Andrade[6], P. P. Khincha[6], S. A. Savage[6], J. Schiffman[7], A. Villani[1], A. Shlien[1], A. Goldenberg[1,2,3,8], D. Malkin[1,2]; [1]The Hospital for Sick Children, Toronto, ON, Canada, [2]University of Toronto, Toronto, ON, Canada, [3]Vector Institute, Toronto, ON, Canada, [4]Royal Children's Hospital Melbourne, Melbourne, Australia, [5]St Jude Children's Research Hospital, Memphis, TN, [6]National Cancer Institute, Bethesda, MD, [7]University of Utah, Salt Lake City, UT, [8]CIFAR: Child and Brain Development, Toronto, ON, Canada, Toronto, ON, Canada
Disclosure Block: **V. Subasri:** None.
Li-Fraumeni syndrome (LFS) is an autosomal dominant cancer-predisposition syndrome associated with pathogenic germline variants in the *TP53* tumour suppressor gene, in approximately 70% of cases. The early identification of genetic susceptibility is especially important in these patients, as they are at markedly increased risk of developing a spectrum of early-onset malignancies. The impact of early identification of germline cancer-causing aberrations in these patients is substantial, as it informs the prospective management of patients and their families. As such, it is imperative to identify cancer-causing aberrations in the remaining ~30% of patients that fit the clinical definition of LFS, but lack a pathogenic germline variant in *TP53*. In addition, even among LFS patients that harbour a pathogenic germline *TP53* variant, the cumulative lifetime risk of developing cancer is approximately 68% in males and 93% in females. The incomplete/variable penetrance, suggests the presence of additional genetic and epigenetic driver events that contribute to increased or decreased cancer risk in certain individuals. Candidate gene approaches have been considered to evaluate other risk factors; however, few studies have utilized unbiased genome-wide DNA sequencing and epigenomic assays to explain this variability in cancer occurrence.

In this work, we leveraged family-based whole-genome sequencing and methylation of DNA procured from peripheral blood leukocytes to evaluate the germline genetic and/or epigenetic genomes of a large cohort of LFS patients (n=396) who harbour either pathogenic variant (n=374) or wildtype *TP53* (n=22). In patients lacking a pathogenic germline variant in *TP53*, we pinpointed the role of alternative cancer-causing genetic aberrations in 8/14 patients that developed cancer. Among *TP53* variant carriers, 19/49 individuals who developed cancer harboured an additional pathogenic variant in another cancer gene, while variants in the WNT signalling pathway were associated with decreased cancer incidence. In addition, we leveraged the non-coding genome and methylome to identify inherited epimutations that confer increased cancer risk. Overall, our study highlights the immense benefits of expanding genetic and epigenetic testing of LFS patients beyond *TP53*, as *TP53* status alone does not indicate whether an individual will develop cancer.

PrgmNr 1414 - Oncogenic mutations in the normal human brain

View session detail

Author Block: **J. Ganz**[1,2,3], E. Maury[1,2,3], B. Becerra[1,2], S. Bizzotto[1,3,2], R. N. Doan[1], C. Kenny[1], T. Shin[1,2,3], J. Kim[1,3,2], Z. Zhou[1,3,2], K. L. Ligon[4,5,2,1,3], A. E. Lee[1,3], C. A. Walsh[1,3,2]; [1]Boston Children's Hosp., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Dana-Farber Cancer Inst., Boston, MA, [5]Brigham & Womenâ□□s Hosp., Boston, MA

Disclosure Block: **J. Ganz:** None.

Oncogenic mutations have been found in non-diseased, proliferative tissues such as skin, blood, and esophagus, showing an age-related increase in most cases. However, the prevalence of such mutations in the normal brain, a low-proliferating organ, is unknown and remains challenging due to the abundance of non-proliferating neurons concentrated in the grey matter. Thus, targeting the white matter for sequencing, which is rich in glial cells, and analyzing data from large cohorts would allow for enhanced power to detect oncogenic events in the non-diseased brain. We evaluated genes implicated in brain tumors by deep sequencing in over 418 normal brain samples derived from 110 individuals of varying ages, without tumor diagnosis, and separately analyzing white and grey matter and other brain regions. We detected predicted and reported pathogenic oncogenic variants in cancer driver genes such as *IDH1, PTPN11, NF1,* and *PTEN*. These mutations were predominantly present in the subcortical white matter (p=0.029) and, surprisingly, were less common in older individuals (p=0.025). In addition, we identified the recurrence of glioma driver variants and their enrichment in glial cells. These findings were replicated using 1,640 non-diseased bulk RNA-seq brain samples from the GTEx consortium, confirming the existence of oncogenic variants and depletion of protein disruptive variants with age. The normal brain exhibits enrichment of mutational signatures present in brain tumors (p=0.00018), suggesting that mutational processes of the normal brain drive early oncogenesis. Our study helps understand the origin and early evolution of brain tumors and opens avenues for early interventions and more accurate diagnosis baselines.

PrgmNr 1415 - A human importin-Ã-related disorder: Syndromic thoracic aortic aneurysm caused by bi-allelic loss-of-function variants in *IPO8*

View session detail

Author Block: **A. Verstraeten**[1], I. Van Gucht[1], J. A. Meester[1], J. Rodrigues Bento[1], M. Bastiaansen[1], J. Bastianen[1], I. Luyckx[1,2], L. Van Den Heuvel[1], P.-J. Guns[3], E. Fransen[1], J. Velchev[1], M. H. A. Perik[1], I. Pintelon[3], M. P. Ferla[4], A. R. Dallosso[5], Genomics England Research Consortium, J. A. Rosenfeld[6], K. Vijayakumar[7], Z. N. Al-Hassnan[8], J. Vogt[9], N. Revencu[10], I. Maystadt[11], E. M. Miller[12], K. Weaver[12], A. H. Begtrup[13], R. Maroofian[14], A. T. Pagnamenta[4], L. Van Laer[1], B. L. Loeys[1]; [1]University of Antwerp/Antwerp University Hospital, Edegem, Belgium, [2]Radboud University Nijmegen Medical Center, Nijmegen, Netherlands, [3]University of Antwerp, Edegem, Belgium, [4]University of Oxford, Oxford, United Kingdom, [5]Southmead Hospital, Bristol, United Kingdom, [6]Baylor College of Medicine, Houston, TX, [7]University Hospitals Bristol NHS Foundation Trust, Bristol, United Kingdom, [8]King Faisal Specialist Hosp, Riyadh, Saudi Arabia, [9]Birmingham Womenâ□□s and Childrenâ□□s Hospital, Birmingham, United Kingdom, [10]University of Louvain, Brussels, Belgium, [11]Institut de Pathologie et de GÃ©nÃ©tique, Gosselies, Belgium, [12]Cincinnati Childrenâ□□s Hospital Medical Center and University of Cincinnati College of Medicine, Cincinnati, OH, [13]GeneDx, Inc., Cincinnati, OH, [14]UCL Queen Square Institute of Neurology, London, United Kingdom

Disclosure Block: **A. Verstraeten:** None.

Importin 8, encoded by *IPO8*, is an ubiquitously expressed member of the importin-Î² protein family that translocates cargo molecules such as proteins, RNAs and ribonucleoprotein complexes into the nucleus in a RanGTP-dependent manner. Current knowledge of the cargoes of importin 8 is limited, but TGF-Î² signaling components such as SMAD1-4 have been suggested to be amongst them. Here, we report that bi-allelic loss-of-function variants in *IPO8* cause a syndromic form of thoracic aortic aneurysm (TAA) with clinical overlap with Loeys-Dietz and Shprintzen-Goldberg syndrome. Seven individuals from six unrelated families showed a consistent phenotype with early-onset TAA, motor developmental delay, connective tissue findings and craniofacial dysmorphic features. A C57Bl/6N *Ipo8* knock-out mouse model recapitulates TAA development from 8-12 weeks onwards in both sexes, but most prominently shows ascending aorta dilatation with a propensity for dissection in males. Compliance assays suggest augmented passive stiffness of the ascending aorta in male *Ipo8*$^{-/-}$ mice throughout life. Immunohistological investigation of mutant aortic walls reveals elastic fiber disorganization and fragmentation along with a signature of increased TGF-Î² signaling, as evidenced by nuclear pSmad2 accumulation. RT-qPCR assays of the aortic wall in male *Ipo8*$^{-/-}$ mice demonstrate decreased *Smad6/7* and increased *Mmp2* and *Ccn2 (Ctgf)* expression, reinforcing a role for dysregulation of the TGF-Î² signaling pathway in TAA development. As importin 8 is the most downstream TGF-Î²-related effector implicated in TAA pathogenesis so far, it offers opportunities for future mechanistic studies and represents a candidate drug target for TAA.

PrgmNr 1416 - Genotype-phenotype correlations in *PIK3CA*-related overgrowth spectrum (PROS) and overlapping phenotypes: a systematic review of 1007 patients with *PIK3CA* pathogenic variants

View session detail

Author Block: **D. Carli**[1], C. Leoni[2], M. Iacoviello[3], A. Mussa[1], C. Ranieri[3], A. Pantaleo[3], P. S. Buonuomo[4], B. Rosanna[3], G. B. Ferrero[5], D. Melis[6], S. Maitz[7], D. C. Loconte[3], A. Turchiano[3], M. Piglionica[3], A. De Luisi[3], F. C. Susca[3], N. Bukvic[3], C. Forleo[8], A. Selicorni[9], G. Zampino[2], G. Cappuccio[10], L. Garavelli[11], C. Novelli[12], L. Memo[13], C. Morando[13], M. Capurso[3], A. Cereda[14], M. Digiacomo[15], V. Saletti[16], A. M. Spinelli[17], P. Lastella[18], V. Dvorakova[19], A. D. Irvine[19], N. Resta[20]; [1]Pediatric Clinical Genetics Unit, Department of Public Health and Pediatric Sciences, University of Torino, and Regina Margherita Childrenâ□□s Hospital, CittÃ della Salute e della Scienza, Torino, Italy, [2]Center for Rare Diseases and Birth Defects, Department of Woman and Child Health and Public Health, Fondazione Policlinico Universitario A. Gemelli IRCCS, Roma, Italy, [3]Department of Biomedical Sciences and Human Oncology (DIMO), Division of Medical Genetics, University of Bari â□□Aldo Moroâ□□, Bari, Italy, [4]Rare Diseases and Medical Genetics Unit, Bambino GesÃ¹ Children's Hospital IRCCS, Roma, Italy, [5]Department of Clinical and Biological Sciences, University of Torino, Torino, Italy, [6]Department of Medicine, Surgery and Dentistry 'Scuola Medica Salernitana', University of Salerno, Salerno, Italy, [7]Clinical Pediatric Genetics Unit, Pediatrics Clinics, MBBM Foundation, S. Gerardo Hospital, Monza, Italy, [8]Cardiology Unit, Department of Emergency and Organ Transplantation, University of Bari Aldo Moro, Bari, Italy, [9]Centro Fondazione Mariani per il Bambino Fragile ASST-Lariana Sant'Anna Hospital, Department of Pediatrics, San Fermo della Battaglia, Como, Italy, [10]Department of Translational Medicine, Section of Paediatrics, Federico II University, Napoli, Italy, [11]Medical Genetics Unit, Mother and Child Health Department, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy, [12]Department of Clinical Sciences and Community Health, The University of Milan, Milano, Italy, [13]Department of Pediatrics, Neonatal Intensive Care Unit, San Bortolo Hospital, Vicenza, Italy, [14]Pediatric Department, ASST Papa Giovanni XXIII, Bergamo, Italy, [15]Unit of Pathology and Medical Genetics, Azienda Ospedaliera Regionale San Carlo, Potenza, Italy, [16]Department of Pediatric Neuroscience, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milano, Italy, [17]Regional Coordinating Center for Rare Diseases, University Hospital, Udine, Italy, [18]Centro Sovraziendale di Assistenza e Ricerca per le Malattie Rare, Internal Medicine Unit 'C. Frugoni', Ospedale Consorziale Policlinico di Bari, Bari, Italy, [19]Dermatology, Children's Health Ireland at Crumlin, Dublin, Ireland, [20]Department of Biomedical Sciences and Human Oncology (DIMO), Division of Medical Genetics, University of Bari â□□Aldo Moro, Bari, Italy

Disclosure Block: **D. Carli:** None.

Purpose: Post-zygotic activating *PIK3CA* variants are responsible for the several phenotypes under the *PIK3CA*-Related Overgrowth Spectrum (PROS). We describe 150 new patients referred for genetic testing, reporting data on their phenotype and underlined *PIK3CA* or *GNAQ, GNA11, RASA1*, and *TEK* variants. A detailed genotype-phenotype correlation in a literature-derived cohort of 1007 *PIK3CA*-mutated patients was also provided. **Methods:** We performed targeted NGS on DNA extracted from blood or buccal swab and tissue biopsy using a custom panel including genes involved in the PI3K/AKT/mTOR pathway and â□□vascularâ□□ genes (*RASA1, TEK, GNAQ,* and *GNA11*). Clinical and molecular data from literature were systematically reviewed and included in the search for correlations. **Results:** *PIK3CA* pathogenetic variants were identified in 93 of 150 unrelated patients. Fifty-seven had a wild type *PIK3CA* allele: pathogenetic variants in *GNA11*, *RASA1, GNAQ*, and *TEK* were found in 11 of them. Differences in the distribution of the variants across *PIK3CA* domains, in the Variant Allele Fraction (VAF) in different tissues, and in the degree of hyperactivation of the PI3K network related to the variant (variant strength) were found compared to current literature. Alone, 10 of the *PIK3CA* variants reported were responsible for more than 70% of cases, including the three

most common mutational hotspots usually reported in several cancers. Combining our cases with those from literature review we draw up a detailed list of all the 81 pathogenic variants described so far in PROS and related them to the respective phenotypes. Eight novel pathogenic variants were also reported. While some *PIK3CA* variants were exclusively associated with a specific PROS phenotype, some were scattered across all the phenotypes and others demonstrated enrichment in some specific phenotypes. Correlations between variant strength and absence of involvement of the central nervous system were also evident. VAF was not correlated with disease severity, and we report severe phenotypes with very low VAF. Patients with pathogenic variants in vascular genes clinically overlapped with PROS. **Conclusion:** Our findings combined with a review of the literature show novel genotype-phenotype correlations underlining the importance of performing a deep phenotyping, carry out a representative tissue sampling, and adopt a comprehensive molecular approach in PROS and overlapping phenotypes.

PrgmNr 1417 - Exome CNV calling and analysis in a large cohort of families with undiagnosed rare genetic disease

View session detail

Author Block: **G. Lemire**[1,2], E. Evangelista[1], I. Wong[1], E. England[1,2], K. Chao[1], L. Pais[1,2], J. Goodrich[1], C. Austin-Tse[1], S. DiTroia[1], E. O'Heir[1,2], J. Fu[1], Broad CMG Collaborators, M. Babadi[1], S. Lee[1], A. N. Smirnov[1], L. Gauthier[1], S. Baxter[1], D. MacArthur[3,1], H. Rehm[1], M. Talkowski[1], H. Brand[1], A. O'Donnell-Luria[1,2]; [1]Broad Institute of MIT and Harvard, Cambridge, MA, [2]Boston Childrenâ□□s Hospital, Boston, MA, [3]Murdoch Childrenâ□□s Research Institute, Melbourne, Australia
Disclosure Block: **G. Lemire:** None.
The Center for Mendelian Genomics (CMG) at the Broad Institute has sequenced 7,719 families with a suspected genetic disease since 2016. Many had a chromosomal microarray and gene panel sequencing for known causes of disease prior to exome sequencing through the CMG. For typical rare variant analysis, exome sequencing data can be used to call SNVs and indels smaller than 50 base pairs, and standard chromosomal microarrays will detect CNVs larger than 50 kilobases. However, mid-sized CNVs are not detected in routine exome analysis of SNVs and indels but molecular diagnostic laboratories are increasingly including CNV calling on exome data in analysis. Detecting CNVs from exome data has been notoriously difficult due in part to the non-uniform distribution of captured reads secondary to biases introduced by PCR and capture steps and many different algorithms have been developed for this purpose. The Genome Analysis Toolkitâ□□s (GATK) CNV tool, the Germline CNV (gCNV) caller, uses a probabilistic framework to infer rare CNVs from read depth data in the presence of systematic bias. We used the gCNV algorithm to call CNVs across the Broad CMG cohort. While analysis is ongoing, we have diagnosed 138 previously unsolved families to date. The identified CNVs consisted of 109 deletions, 20 duplications and 9 complex CNVs. A CNV in a known gene that is consistent with the phenotype was identified in 114 families and a CNV in a novel candidate gene was identified in 24 families. Supporting genetic and/or experimental evidence were required to consider a CNV in a novel gene as the diagnosis in a given family, most often by additional families identified through Matchmaker Exchange. The predominant phenotype present in these families were neurodevelopmental disorders (67%) followed by neuromuscular disorders (19%). We estimate that about 50% of the CNVs that solved CMG cases would not have been detected by standard chromosomal microarrays. Calling CNVs from existing exome data increases the diagnostic yield for individuals that remain undiagnosed after standard testing approaches, providing a higher resolution alternative to arrays at a fraction of the cost of genome sequencing.

PrgmNr 1418 - Brazil enters the genome sequencing era: The Rare Genomes Project - interim results

View session detail

Author Block: R. M. Minillo[1], E. Perrone[1], C. R. D. Quaio[1], L. A. Virmond[1], K. Chen[2,1], C. A. Moreno[1], J. M. Ceroni[1], T. F. Almeida[1], A. V. C. Coelho[1], N. B. Zurro[1], G. B. C. Colichio[1], R. S. Reis[1], M. C. Cervato[1], M. P. Caraciolo[1], R. B. Alexandre[1], F. C. Val[1], G. S. Oliveira[1], D. D. C. Oliveira[1], L. M. S. Moura[1], P. N. Chazanas[1], M. Gretschischkin[1], B. M. C. Azevedo[1], M. S. Nobrega[1], C. L. P. Mangueira[1], **J. O. Filho**[1]; [1]Albert Einstein Hosp., SÃ£o Paulo, Brazil, [2]SÃ£o Paulo, Brazil

Disclosure Block: **J.O. Filho:** None.

Rare diseases (RD) are estimated to affect between 3.2 and 13.2 million individuals in Brazil. The Rare Genomes Project (RGP) is a Brazilian initiative for whole genome sequencing (WGS) of patients with RD recruited from centers for rare diseases of the National Unified Health System (SUS). RGP initiated in 2020 and will sequence the genomes of 7755 patients with RD until the end of 2023, allowing the creation of the largest Brazilian genomic database of patients with RD and hereditary cancer. RGP also aims to study the clinical and social burden of RD in Brazil, provide valuable information for cost-effective national health policies for a population generally underrepresented in previous genomic studies and improve diagnosis, therapy, prevention, genetic counseling and the quality of life of patients. The ongoing project has recruited 1486 probands from nine centers of four out of five macroregions of the country. Among those, approximately 1300 have already had their genomes sequenced, and for 163 clinical reports were returned to attending physicians.The most prevalent disorders are neurological (n=402, average age:11.8 years), congenital malformations (n=376, average age:10.6 years), hereditary cancer (n=343; average age:45), inborn errors of immunity (n=145; average age:19.6 years) and clinical genetic syndromes (n=114; average age:10.9 years). Out of the 163 reports released, molecular diagnosis was established for 57 patients (34,9%). In 43 patients, no candidate variants were detected (26,38%). The remaining 63 (38,7%)patients presented variants of unknown significance (VUS) in heterozygous or homozygous state, heterozygous state for a recessive phenotype or pathogenic/likely pathogenic (P/LP) variants in heterozygous state and a VUS in the same gene for a recessive phenotype. Among the positive cases, the higher diagnostic rates were seen in the neurological (18 cases; 36%), clinical genetic syndromes (10 cases; 17.6%), inborn errors of metabolism (7; 12.3%), immunodeficiencies (5; 8.8%), and hereditary cancer syndromes (5;8.8%) cohorts. A total of 81 variants were detected in the 57 positive patients (67 SNVs and 14 large-scale CNVs/SVs), 69 (85.2%) being annotated as (P/LP) (57 SNVs and 12 CNVs/SVs). Interesting to mention that 112 positive cases (21.1%) had P/LP CNVs or SNVs in regulatory regions which would not be identified by whole exome sequencing (WES). In summary, the RGP is the largest rare disease sequencing effort in Brazil, and the preliminary results demonstrate a high diagnostic rate.

PrgmNr 1419 - Low-pass genome sequencing-based detection of absence of heterozygosity: validation in clinical cytogenetics

View session detail

Author Block: **Z. Dong**[1], M. Chau[1], Y. Zhang[1], Z. Yang[1], m. Shi[2], Y. Wah[1], Y. Kwok[1], T. Leung[3], C. C. Morton[4], K. choy; [1]The Chinese University of Hong Kong, Hong Kong, Hong Kong, [2]Hong Kong, China, [3]The Chinese University of Hong Kong, Hong Kong, China, [4]Brigham & Women's Hosp, Boston, MA
Disclosure Block: **Z. Dong:** None.
Introduction: Absence of heterozygosity (AOH) is a genetic characteristic known to cause human genetic disorders through autosomal recessive or imprinting mechanisms. However, the analysis of AOH via low-pass genome sequencing (GS) is not yet clinically available. **Materials and Methods:** Low-pass GS (fourfold) with different types of GS libraries was performed on 17 clinical samples with previously ascertained AOH by chromosomal microarray analysis (CMA). In addition, AOH detection was performed with low-pass GS data in 1,639 cases that had both GS and high-probe density CMA data available from the 1000 Genomes Project. Cases with multiple AOHs (coefficient of inbreeding F â¥ 1/32) or terminal AOHs â¥5 Mb (suspected uniparental disomy [UPD]) were reported based on the guidelines of the American College of Medical Genetics and Genomics. **Results: We first demonstrated the optimal read-depth for AOH analysis to be fourfold regardless of sequencing modes (paired-end or single-end) and library types (small-insert or large-insert). In addition,** low-pass GS revealed suspected segmental UPD and multiple AOHs (F â¥ 1/32) in nine and eight clinical cases, respectively, consistent with CMA. Lastly, among the 1,639 samples with CMA and GS available in the 1000 Genomes Project, low-pass GS not only consistently detected multiple AOHs (F â¥ 1/32) in 18 cases, but also reported 60 terminal AOHs in 44 cases including four mosaic AOHs at a level ranging from 50% to 75%. **Conclusion:** Overall, our study demonstrates the feasibility of AOH analysis (â¥5 Mb) with low-pass GS data and shows high concordance compared with CMA.

PrgmNr 1420 - Advanced Diagnostics and Genotype-Phenotype Resolution using Functional Genomics in >500 Neuromuscular and Neurological Disorder Patients

View session detail

Author Block: **S. Chakravorty**[1], K. Berger[2], L. Rufibach[3], S. Shira[3], S. Verma[1], R. Logan[1], M. Wicklund[4], M. B. Harms[5], T. Mozaffar[6], V. E. Kimonis[7], D. Arafat[2], G. C. Gibson[8], M. R. Hegde[9]; [1]Emory Univ., Atlanta, GA, [2]Georgia Inst. of Technology, Atlanta, GA, [3]Jain Fndn. Inc., Seattle, WA, [4]Univ. of Colorado Neurology, Denver, CO, [5]Washington Univ. Sch. of Med., St. Louis, MO, [6]Univ. of California Irvine Neurology, Irvine, CA, [7]Univ CA Irvine, Orange, CA, [8]Georgia Tech, Atlanta, GA, [9]PerkinElmer, Lilburn, GA

Disclosure Block: **S. Chakravorty:** None.

Background: 50-70% of inherited rare neuromuscular disease (NMD) patients remain undiagnosed even after DNA testing, a barrier for clinical trial enrolment. Recently, using a muscular-dystrophy next-generation-sequencing panel on 4656 congenital/limb-girdle muscular dystrophy (CMD/LGMD)-suspected patients across the US and using exome sequencing on 207 genetic myopathies across the Indian subcontinent, and investigating hereditary NMDs and peripheral neuropathies at CHOA, we identified the major hurdles were: a) lack of genotype-phenotype correlation, b) high prevalence (72%) of variants of uncertain significance (VUSs), c) >30% of all patients had pathogenic variant(s) or VUSs in â¥2 genes (multi-genic), and d) the lack of less-invasive biomarker-driven approaches. **Methods:** Here, we used high-throughput clinical-grade RNA-Seq with a proprietary tiered analytical pipeline, and co-immunoprecipitation combined with mass-spectrometry proteomics, and other targeted assays on muscle/skin/blood biopsy-tissues to resolve VUSs and multi-genic cases to enhance molecular diagnostics, and to resolve genotype-phenotype relationships. **Results:** Using targeted RNA-Seq analysis and other assays with genotype-clinical-data correlation on 548 cases, we achieved 64% diagnostic yield and 88% diagnostically informative results. Besides VUS reclassification, we identified variant mechanisms acting either by abnormal splicing/allele/gene-expression/protein-stability/function levels or causing defects in pathways. For example, an 8-year old child with proximal weakness, dystrophic changes on muscle biopsy with normal immunohistochemistry including alpha sarcoglycan stains harbored two variants (pathogenic: c.229C>T, likely pathogenic: c.957-11C>G) *in trans* in *SGCA*. This genotype-muscle biopsy discrepancy was resolved using above-mentioned technique by identifying normal mRNA and protein expression for all Sarcolgycans in muscle even with c.957-11C>G causing both exons 6-7 and 6-8 skips, and that the variantsâ pathogenicity acting at protein function level. Furthermore, the novel application helped discover a new gene, *DRGX*, associated with bilateral hand weakness, finger flexor contractures and sensory motor polyneuropathy in a teenager. Additionally, using enzyme assays and RNA-Seq, we reclassified 20 *GAA* (Î±-glucosidase) gene VUSs as pathogenic variants to resolve undiagnosed Pompe disease cases. **Conclusions:** Our results show the importance of using a multi-tiered approach that includes omics platforms, biomarkers and genotype-phenotype correlation not only for diagnostics but also for better trial-readiness.

PrgmNr 1421 - Functional characterization of haplotype surrounding *TOMM40*-523â repeat to assess differential risk effects on European ancestry *APOEÉ3* haplotypes

View session detail

Author Block: **M. Lipkin Vasquez**, P. Bussies, F. RAJABLI, K. L. Hamilton-Nelson, A. J. Griswold, M. A. Pericak-Vance, J. Young, J. M. Vance, K. Nuytemans; John P. Hussman Institute for Human Genomics, Miami, FL
Disclosure Block: **M. Lipkin Vasquez:** None.
Introduction: Reports on involvement of *TOMM40*, a gene neighboring *APOE*, in Alzheimer Disease (AD) risk have been inconsistent. Recently, we showed that the length of a poly-T repeat in *TOMM40* (*TOMM40*-523â) is associated with AD risk in individuals carrying *APOEÉ3*, the most common *APOE* haplotype in the general population, on European local ancestry (LA). âVery longâ repeat lengths (VL, >29T) have a protective effect compared to âshortâ repeat lengths (S, APOE*É3* or *APOEÉ4* haplotypes in either ancestral background. Therefore, we hypothesized that variants in linkage disequilibrium (LD) with *TOMM40*-523â on the European LA *APOEÉ3* haplotype can modify risk for AD, potentially through *APOE* regulation. Methods: We used the short tandem repeat detection bioinformatics algorithm HipSTR to type S and VL repeats in whole genome sequencing data of individuals homozygous for the *APOEÉ3* European LA haplotype from the Puerto Rico AD Initiative (PRADI) project. Frequency of variants on 16 S and 14 VL independent haplotypes were compared to determine variants in LD with the repeat. HaploView was used to determine the LD structure of the repeat and surrounding variants. Results: We identified a 16kb LD block surrounding *TOMM40*-523â harboring 21 variants in strong LD (r^2>0.9) with the repeat (hg19, chr19:45,395-45,411k). This region includes the putative *APOE* promoter (harboring LD variants rs405509, rs449446 and rs769450) and a *TOMM40* intronic region with previously reported enhancer activity (harboring LD variants rs157580, rs2075649 and rs157584). Assessment of combined regulatory function of variants in LD with the repeat on S or VL haplotypes in the *APOE* promoter and *TOMM40* enhancer region is currently ongoing using luciferase reporter assays in AD-relevant cell types (i.e. astrocytes, microglia and neurons). Discussion: The identification of clearly distinct S and VL haplotypes on *APOEÉ3* European LA background support importance of the surrounding variants in the risk effect observed in the association analyses. The LD analyses data suggest that *TOMM40*-523â LD variants could directly affect *APOE* expression through presence in the promoter itself and/or in an identified enhancer region in *TOMM40*. The follow-up functional data will pinpoint the driving regulatory element(s) and *TOMM40*-523â LD variants for the observed different AD risk effects in European ancestry *APOEÉ3* carriers. Long term, treatments targeting these regulatory regions may be relevant to a large amount of people given *APOEÉ3*âs frequency in the general population.

PrgmNr 1422 - Telomere length analysis in amyotrophic lateral sclerosis using large-scale whole genome sequence data

View session detail

Author Block: **A. Al Khleifat**, The Project MinE Consortium; Kingâ□□s College London, London, United Kingdom
Disclosure Block: **A. Al Khleifat:** None.
Background Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease characterized by the loss of upper and lower motor neurons, leading to progressive weakness of voluntary muscles, with death following from neuromuscular respiratory failure, typically within 3 to 5 years. There is a strong genetic contribution to ALS risk. In 10% of cases or more, a family history of ALS or frontotemporal dementia is obtained, and the Mendelian genes responsible for ALS in such families have now been identified in about 80% of cases. Only about 14% of apparently sporadic ALS is explained by known genetic variation, suggesting that other forms of genetic variation are important. Telomeres maintain DNA integrity during cellular replication, differ between sexes, and shorten naturally with age. Gender and age are risk factors for ALS and we therefore investigated telomere length in ALS.
Methods: Samples were from Project MinE, an international ALS whole genome sequencing consortium that includes phenotype data. For validation we used donated brain samples from motor cortex from people with ALS and controls. Ancestry and relatedness were evaluated by principal components analysis and relationship matrices of DNA microarray data. Whole genome sequence data were from Illumina HiSeq platforms and aligned using the Isaac pipeline. We estimated telomere length by applying a bioinformatics analysis to the data. We tested the association of telomere length with ALS and ALS survival. **Findings**: There were 6,580 whole genome sequences, reducing to 6,195 samples (4,315 from people with ALS and 1,880 controls) after quality control, and 159 brain samples (106 ALS, 53 controls). Accounting for age and sex, there was a 20% (95% CI 14%, 25%) increase of telomere length in people with ALS compared to controls ($p = 1.1 \times 10^{-12}$), validated in the brain samples ($p = 0.03$). Those with shorter telomeres had a 10% increase in median survival ($p = 5.0 \times 10^{-7}$). Although there was no difference in telomere length between sporadic ALS and familial ALS ($p=0.64$), telomere length in 382 people with ALS due to expanded *C9orf72* repeats was less than in those without expanded *C9orf72* repeats ($p = 5.0 \times 10^{-4}$). **Interpretation:** Although telomeres shorten with age, longer telomeres are a risk factor for ALS and worsen prognosis.
Conclusions: It is likely that longer telomeres increase risk for ALS.

PrgmNr 1423 - Shared genetic aetiology of type 2 diabetes and knee osteoarthritis

View session detail

Author Block: **A. Arruda**[1], E. Zeggini[2,3]; [1]Institute of Translational Genomics, Helmholtz Association Munich, Neuherberg, Germany, [2]Helmholtz Association Munich, Neuherberg, Germany, [3]Technical University Munich and Klinikum Rechts der Isar, Munich, Germany
Disclosure Block: **A. Arruda:** None.
Type 2 diabetes (T2D) and knee osteoarthritis (OA) are two of the most prevalent chronic diseases worldwide. Both disorders share common risk factors, for instance, obesity and increasing age, and frequently coexist in older adults. Observational studies report a positive epidemiological association between the diseases beyond their common risk factors. Mendelian randomization has shown no causal relation between T2D and knee OA, whereas body-mass index (BMI) is a causal factor for both. Taking into consideration the increase of the worldâ□□s elderly population and the chronic nature of these comorbid diseases, understanding their shared aetiology is of utmost importance.
Using summary statistics of large-scale genome-wide association studies (GWAS) from T2D (n=898,130) and knee OA (n=490,345), we investigate the genetic intersection between the traits by performing statistical co-localization analysis of established association signals using *COLOC*. For co-localizing signals, we then perform a multi-trait co-localization analysis using *HyPrColoc* with eQTL information derived from four disease-specific human tissues: pancreatic islets, intact OA cartilage, degenerated OA cartilage and OA synovium. We further integrate eQTL information from GTEx.
We find robust evidence for co-localization of T2D and knee OA signals in multiple regions. For instance, a region around the *FTO* gene co-localizes with a posterior probability of a shared causal variant (PP4) of 0.9. The 95% credible set consists of six variants, with *rs1421085* as the lead SNP. *rs1421085* (C) is associated with increased BMI. Moreover, rs*1421085* is an eQTL for *IRX3* in the synovium tissue and pancreatic islets, and these eQTLs co-localize with the T2D and knee OA signals. The associated *FTO* variants are known to form a long-range functional connection to *IRX3*, interacting with its promoter and regulating its expression. Our results underline the shared epidemiological association of knee OA and T2D with obesity and BMI, by showing co-localization of both traits and an association to disease-relevant tissues in the *FTO* obesity-related region.

PrgmNr 1424 - Genetic control of mRNA splicing is affected by purifying selection and may be linked to incomplete penetrance

View session detail

Author Block: **J. Einson**[1,2], T. Lappalainen[2], D. A. Glinos[2]; [1]Columbia University, New York, NY, [2]New York Genome Center, New York, NY
Disclosure Block: **J. Einson:** None.
Common variants affecting mRNA splicing are typically identified by splicing quantitative trait locus (sQTL) mapping and have been shown to be enriched for GWAS signals by a similar degree to eQTLs. However, the specific splicing changes induced by these variants have been difficult to characterize, making it more complicated to analyze the effect size and direction of sQTLs.Furthermore, sQTLs may affect the dosage of LoF variants in their target exons. This scenario is a potential driver of incomplete penetrance. To test our model, we first catalogued sQTLs using RNA-seq and WGS data from GTEx v8, using each exonâ□□s percentage spliced in (PSI) metric as a quantitative phenotype. PSI is an interpretable way of assessing an sQTLâ□□s effect size and direction. In total, we identify 5,196 genes with at least one significant exon across at least 1 of 18 GTEx tissues. With this set of sQTLs, it is more common that the derived alleles decrease (n=2,744) rather than increase (n=2,185) the inclusion of their target exons, but have a lower allele frequency distribution compared to sQTLs that increase exon inclusion (K.S. test $p=1.998Ã□10^{-15}$). This suggests purifying selection is acting on sQTL variants based on their regulatory properties.We also performed colocalization analysis between sQTL and GWAS loci across 18 tissues and 114 GWAS studies. We found many examples of sQTL variants colocalizing with GWAS hits, and found some evidence that sQTL effect size, direction and other properties influence the likelihood of a significant colocalization event. Finally, we tested whether sQTLs modifying inclusion of their target exons may modify the penetrance of rare coding variants on the same haplotype. To this end, we analyzed signs of purifying selection by looking for depletion of high penetrance haplotype configurations in a general population. We tested for depletion of high penetrance haplotypes first using all 838 individuals in GTEx v8 with WGS, and then in the larger Trans-Omics for Precision Medicine (TOPMed) project, which includes whole genome sequencing from 63,420 individuals of European descent. The larger sample size allows us to probe penetrance patterns in ultra-rare variants, more common variants, and in genes with multiple rare variants across individuals. Ultimately, we provide insights into the multiple mechanisms how genetic effects on splicing contribute to patterns of genetic variation in human populations and genetic disease risk for common and rare diseases. This technique could improve our interpretation of the risks associated with genetic variation beyond the exome.

PrgmNr 1425 - Rare variants affecting multi-omic measurements of gene regulation implicated in EKG traits

View session detail

Author Block: **T. Li**[1], M. Arvanitis[1], B. J. Strober[1], B. Ni[1], V. Wang[1], J. K. Bonnie[1], R. Keener[1], H. J. Lin[2], J. I. Rotter[2], J. A. Brody[3], S. R. Heckbert[3], N. Sotoodehnia[3], D. E. Arking[4], A. Battle[1], NHLBI TOPMed consortium; [1]Johns Hopkins University, Baltimore, MD, [2]Lundquist Institute, Harbor-UCLA Med Ctr, Torrance, CA, [3]University of Washington, Seattle, WA, [4]Johns Hopkins School of Medicine, Baltimore, MD

Disclosure Block: **T. Li:** None.

Biobank-scale whole genome sequencing has identified molecular pathways relevant to disease that are disrupted by rare coding variants. However, interpretation of non-coding rare variants (RV) is still challenging. RVs have been implicated in complex diseases such as atrial fibrillation (AF). AF variants can be analyzed with electrocardiogram (EKG) traits, but mechanisms of associations remain elusive. Here, we leveraged transcriptomic, proteomic, and methylomic data from Multi-Ethnic Study of Atherosclerosis (MESA) in the TOPMed consortium, to evaluate functional RVs that affect EKG traits through expression of nearby genes in each omic signal. For a subset of 1816 European ancestry individuals, we calculated multi-omic z-scores after regressing out confounders. We observed enrichment of RVs near RNA, protein, and methylation signals with outlier z-scores. We used a hierarchical Bayesian model, Watershed, to prioritize functional RVs and assign posterior probabilities of functional effects for each triplet (gene, individual, RV; 104M analyzed). Compared to WGS annotations, our model significantly improved prediction of multi-omic outlier status for individuals with shared RVs (AUROC = 0.65). The model learned genomic features that predict outlier expression (e.g. TF binding for RNA, frameshift variants for protein). There was good correlation between z-scores of nearby genes and corresponding posteriors for RVs shared with GTEx. Watershed improved risk stratification for diverse polygenic traits over common variants. To study contributions of RVs to EKG traits, we built a framework to evaluate the excess of functional RV burden per gene on 5 EKG traits representing cardiac depolarization and conduction (PR, QRS); repolarization (QT, JT); and rate (RR interval). Multi-omic posteriors prioritized complementary genes associated with extreme EKG values. Extreme QRS intervals were associated with protein outliers of *CFC1*, a risk gene for congenital heart disease, and with mRNA outliers of *STUB1*, a ubiquitin ligase in the heart and striated muscle. A Cox proportional-hazards model showed that extreme values in all EKG traits, except JT, were significantly associated with poorer survival. In a mediation model, 40 genes had outlier protein levels potentially associated with poorer survival through extreme QRS values - for example *ARCN1*, part of the coatomer complex involved in developmental disorders. Our comprehensive survey of RVs and multi-omic outlier signals in a European cohort yields mechanistic insights into EKG traits. These methods are a framework for prioritizing functional RVs in precision medicine studies.

PrgmNr 1426 - Systematic evaluation and Improvements for trans-eQTL Detection Methods Allows Identification of Novel trans-eQTLs in the GTEx data

View session detail

Author Block: **C. Wu**, T. Xu, A. Goren, M. Gymrek; University of California, San Diego, San Diego, CA
Disclosure Block: **C. Wu:** None.
Studies of expression quantitative trait loci (eQTLs) have aimed to discover genetic variants that explain variation in gene expression levels due to associations with complex traits and human diseases. While thousands of cis-eQTLs have been reliably identified, consistently replicating trans-eQTL effects proved to be challenging due to insufficient statistical power, lack of comparable tissues and cohorts, and putative false positive associations. In particular, technical covariates lead to a substantial variation in expression datasets and result in multiple false positive eQTL calls. As a result, biological mechanisms and characteristics of trans-eQTLs remain largely unknown.
Here we present x-qtl â a novel trans-eQTL detection method. For a given trans-eQTL, we assume the distribution of association statistics to be a mixture of two distributions â for the target and non-target genes. x-qtl addresses two critical limitations of existing methods. First, the method allows downstream characterization of trans-eQTLs by predicting the number of target genes. Second, x-qtl applies principal component analysis to the expression matrix to account for gene correlations. Specifically, we represent the variables (genes) as the uncorrelated principal components which reduces noise from correlated genes that can obfuscate true trans-eQTL signals.
Next, we develop a simulation framework to evaluate trans-eQTL detection tools. Our framework enables for varied effect size distribution and number of target genes for each trans-eQTL, includes pairwise correlation between expression of different genes, and simulates effects of technical covariates. We use our framework to evaluate the power of x-qtl and two existing eQTL frameworks, MatrixeQTL and CPMA, to detect trans-eQTLs with a variety of characteristics. Our results indicate that only trans-eQTL with extremely large effect sizes, or affecting hundreds to thousands of target genes, can be reliably detected. Remarkably, x-qtl is more sensitive in detecting trans-eQTLs with a small number of targets than existing methods.
We benchmark x-qtl against existing tools using yeast expression data containing documented eQTLs. Then, we apply all methods to the recently released Genotype-Tissue Expression (GTEx) project v8 dataset with 838 human donors and 15,201 samples from 52 tissues. We observe a putative trans-eQTL in Nerve-Tibial involved in signal transduction pathways and another in Thyroid that is located in a thyroid-specific transcription factor. Additionally, we detect novel trans-eQTLs that are replicated among different tissues but were missed by traditional trans-eQTL analysis methods.

PrgmNr 1427 - Brain eQTL of East Asian, African American, and European Descent Explains Schizophrenia GWAS in Diverse Populations

View session detail

Author Block: **Y. Chen**[1], S. Liu[2], F. Wang[1], Y. Jiang[3], Y. Xia[4], W. Qiu[5], C. Ma[5], X. Yan[1], J. Huang[1], S. Xu[6], B. Tang[1], H. Huang[7], C. LIU[8], C. Chen[1]; [1]Central South University, changsha, China, [2]Central South University, China, China, [3]Nashville, [4]Broad Institute, Medford, MA, [5]Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [6]CAS-MPG Partner Institute for Computational Biology, Shanghai, China, [7]Boston, MA, [8]SUNY Upstate Medical University, Syracuse, NY
Disclosure Block: **Y. Chen:** None.
Previous studies integrating genome-wide association studies (GWAS) and brain expression quantitative trait loci (eQTL) data have discovered hundreds of risk genes associated with schizophrenia (SCZ). However, most studies focused on individuals of European Populations. Based on the differences in population structures such as linkage disequilibrium patterns and allele frequencies, we hypothesized that different populations had their own SNPs associated with brain gene expression; and population-relevant brain eQTL will improve interpretation of GWAS signals of the corresponding populations than the use of eQTL from other populations in SCZ.We first characterized the eQTL using RNA-seq and WGS data of brain samples from subjects of East Asian (EA, N = 228), Africa American (AFR, N = 175), and European (EU, N = 407) ancestries. Using FDR q-value Population shared (PS-), defined as the eQTL, of the identical SNP-gene pair, with the same effect in all three populations; or *Population-unique* (PU-), defined as the eQTL is only significant or has significantly different effect size in one population. Striking differences existed in eQTL across populations: only 3,999 genes were significantly associated with the same SNPs in three populations. We identified 1,308 and 333 PU-eQTL Genes in EA and AFR, respectively. Meanwhile, the SNP minor allele frequency and fixation index values of eQTLs are significantly higher in the PU-eQTL SNP than the PS-eQTL SNP (adjusted p

PrgmNr 1428 - Genome-wide inter-chromosomal epistatic associations identified across complex diseases in the ~300,000 participants from eMERGE and UK Biobank

View session detail

Author Block: **S. S. Verma**[1], P. Singhal[2,3], A. Lucas[4,3], Y. Veturi[5,3], C. Weng[6], s. pendergrass, I. J. Kullo[7], S. J. Schrodi[8,9], D. Fasel[10], D. J. Schaid[7], O. Dikilitas[7,11], P. M. Sleiman[12], H. Hakonarson[13], M. D. Ritchie[1]; [1]University of Pennsylvania, Philadelphia, PA, [2]Philadelphia, MA, [3]University of Pennsylvania, Philadelphia, PA, [4]411 Waupelani Drive A-341, Philadelphia, PA, [5]State College, PA, [6]Columbia University, New York, NY, [7]Mayo Clinic, Rochester, MN, [8]Marshfield Clinical Res Fndn, Marshfield, WI, [9]University of Wisconsin, Madison, WA, [10]New York, NY, [11]Children's Hospital of Philadelphia, Philadelphia, PA, [12]CHOP, Philadelphia, PA, [13]Children S Hospital of Philadelphia, Philadelphia, PA
Disclosure Block: **S.S. Verma:** None.

Capitalizing on linkage disequilibrium (LD) patterns to determine population substructure allows the discovery of additive association signals in genome-wide association studies (GWAS). As standard GWAS analyses are well-powered to interrogate additive models, investigating other plausible modes of inheritance such as dominance and epistasis may often require new approaches. Epistasis - interaction between genes, play an essential role in elucidating complex genetic networks that impact the human genome's structure and evolution. Given that evidence from model organism studies indicates long-range high LD regions to be under evolutionary selection, we hypothesized that these regions might play a key role in regulating disease mechanisms across various complex traits. Thus, we selected 20 diverse complex traits (neurological, ocular, cardiometabolic, immune) and created case/control cohorts of individuals in the eMERGE and UKBB datasets. We calculated Ohtaâ□□s D statistic on genome-wide pairs of variants to identify pairs in long-range (> 0.25cM) LD due to epistatic selection. On these resulting 136,019 SNPs culminating in a total of 5,625,845 SNP-SNP models, we used a penalized regression framework to determine the association of SNP-SNP pairs with a disease. Across 12 of the 20 phenotypes, we found 290 models (majority inter-chromosomal) that replicated in UKBB and eMERGE after multiple hypothesis testing correction. Characterization of replicating models indicate the genes they map to are 1) highly conserved gene families with complex roles in multiple pathways, 2) essential genes, and/or 3) already associated in the literature with complex traits that have diverse phenotypic presentations. Together, these results demonstrate the highly pleiotropic nature of variants under epistatic selection in conserved regions, supporting the hypothesis that epistatic interactions regulate diverse clinical mechanisms and can produce a range of phenotypic outcomes. A key example is the WFS1 gene, coding for the calcium-regulating wolframin protein. Over 200 mutations in WFS1 have been associated with Wolframin Syndrome, characterized by different combinations of diabetes, optic atrophy, urinary tract dysfunction, loss of hearing, and neurological symptoms. Our results support the conclusion that epistatic interactions between *WFS1* and other genes, including *VRK2, NT5C2, INA, and DIP2C*, regulate this syndrome and produce the range of phenotypes seen. This study examines numerous such cases and sheds light on the pleiotropic interactions underlying disease etiologies providing a more unified view of epistasis and pleiotropy in complex traits.

PrgmNr 1429 - Global Biobank Meta-Analysis Initiative: A genome-wide association meta-analysis identifies novel primary open-angle glaucoma loci and shared biology with vascular mechanisms

View session detail

Author Block: **V. Lo Faro**[1], J. Hirbo[2,3], A. Bhattacharya[4], W. Zhou[5,6,7], Global Biobank Meta-analysis Initiative, N. Jansonius[8], H. Snieder[9], K. Joos[10], N. Cox[2,3]; [1]Department of Ophthalmology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands, [2]Division of Genetic Medicine, Vanderbilt University School of Medicine, Nashville, TN, [3]Vanderbilt Genetics Institute, Nashville, TN, [4]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, [5]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, [6]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, [7]Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, [8]Department of Ophthalmology, University Medical Center Groningen, Groningen, Netherlands, [9]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands, [10]Vanderbilt Eye Institute, Vanderbilt University School of Medicine, Nashville, TN

Disclosure Block: **V. Lo Faro:** None.

Primary open-angle glaucoma (POAG) is a complex eye disease characterized by progressive loss of optic nerve function that if untreated ultimately leads to irreversible blindness. To date, the biological mechanisms causing POAG are still unclear. A vascular hypothesis of unstable ocular perfusion has been suggested to explain the process of optic nerve damage. However, no study has previously provided a detailed exploration of the biology underlying the potential vascular connection with POAG. The Global Biobank Meta-Initiative (GBMI), a collaboration of 20 global biobanks, provides an exceptional resource to examine potential shared vascular POAG biology. A large-scale Genome-wide association studies meta-analysis was conducted in subjects sourced from 15 global biobanks and from six ancestries (n=1,487,447). A total of 59 statistically significant loci-trait associations were identified, seven of which were novel. Four loci encompassing the genes *ZFP91-CNTF, GLYAT, KALRN, CCDC13, MIR2054* and *INTU* were tested and replicated in an independent POAG cohort (n=383,500). To add biological context to these variant-trait associations, we performed TWAS using JTI cis models in 24 GTEx tissues potentially relevant to ocular conditions. We then performed fine-mapping identifying 29 gene-trait associations, five of which have been implicated or have vascular related functions: *CDKN2B, SLC35E2A, ITGB5* and *MYL4*. Further, we performed a gene enrichment analysis in which morphology and development of blood vessels, and angiogenesis pathways were significantly enriched. A gene prioritization analysis found 60 co-regulated genes of which 39 were novel. In total, 15 of the 39 novel POAG genes identified were associated with blood regulation, cardiac disease and arterial stiffness measurement. We did extensive statistical validation analysis of genes in *SIX6* and *CDKN2B-AS1* loci, previously implicated in POAG, cardiovascular diseases and cancers across multiple ancestries. Results from this analysis confirmed that the TWAS association signals in these loci are attributed to the sentinel rs33912345 missense variant in the SIX6 gene and variants linked to the *CDKN2B-AS1* gene. We also found evidence of significant interaction between the rs33912345 and causal variants in chr9p21.3, with concomitant effect on expression of the genes *CDKN2A* and *CDKN2B*. We confirmed shared biology between cardiovascular diseases and POAG by performing meta-analysis cis model TWAS-PheWAS across the whole phenome in BioVU and UKbiobank data (n=456,423). Taken together, these findings enforce the contribution of genes involved in vascular mechanisms to POAG pathogenesis.

PrgmNr 1430 - Systematic comparison of family history and polygenic risk across 27 common diseases

View session detail

Author Block: **N. Mars**[1], J. V. Lindbohm[2,3], P. della Briotta Parolo[1], E. WidÃ©n[1], J. Kaprio[1,2], A. Palotie[1,4,5], FinnGen, S. Ripatti[1,6]; [1]Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Helsinki, Finland, [2]Clinicum, Department of Public Health, University of Helsinki, Helsinki, Finland, [3]Department of Epidemiology and Public Health, University College London, London, United Kingdom, [4]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, [5]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, [6]Broad Institute of MIT and Harvard, Cambridge, MA
Disclosure Block: **N. Mars:** None.
Family history is the standard measure of inherited disease susceptibility in clinical practice, while polygenic risk scores (PRS) have recently been shown to capture efficiently personal genetic risks in many diseases. No studies have systematically compared how they overlap and complement each other in a wide range of common complex diseases.
Here we leverage family relationships, information about parental causes of death, and genome-wide genotyping of 306,418 individuals within the FinnGen study, to systematically examine the interplay of first and second-degree family history of diseases, parental causes of death, and genome-wide PRSs. With up to 50 years of follow-up within nationwide health registries, we study 27 diseases, covering a large proportion of the burden of non-communicable diseases in adults. The 27 genome-wide PRSs were constructed using PRS-CS.
The effects of PRS and family history were largely independent and provided complementary information for risk prediction. Both high PRS and family history were strongly associated with respective diseases, but a high PRS (>90th percentile) conferred larger risks than family history in cardiometabolic diseases and in the three common cancers studied. Among 39,444 individuals with a first-degree relative in the dataset, the impact of adjusting the PRS with first-degree family history was small (mean decrease in betas -3.4%, s.d. 1.8%). Similarly, the decrease was small adjusting the effect of first-degree family history with the PRS (-4.9%, s.d. 3.3%). The decreases were even smaller for second-degree family history (N=47,154), and for parental causes of death (N=227,982). In 23/27 diseases, PRS improved risk discrimination beyond first-degree family history, with largest increases in rheumatoid arthritis, asthma, and prostate and breast cancer.
PRSs were effective also for risk stratification among individuals with positive family history. In most diseases, a positive family history with a high PRS was associated with a considerably elevated risk, whereas a low PRS compensated completely for the risk implied by a positive family history. For instance, in type 2 diabetes, individuals with positive family history and high (>90th percentile), average (33-90th) or low (This study provides a catalogue of risk estimates and prediction accuracy for both family history of disease and PRSs. We demonstrate that these metrics are largely independent and complementary measures of familial susceptibility.

PrgmNr 1431 - Systematic comparison of performance of over 50 PRS across ancestries for 14 diseases

View session detail

Author Block: **V. Plagnol**, S. Selzam, W. Tarran, R. Moore, D. Wells, F. Riveros-Mckay, P. Donnelly, G. McVean, M. E. Weale; Genomics plc, Oxford, United Kingdom
Disclosure Block: **V. Plagnol:** Salary/Employment; Genomics plc.
There has been intense recent interest in the performance of PRSs across ancestry groups, in particular in the extent of performance deterioration in non-European individuals. This is particularly pressing as the technology moves closer to use in healthcare.
However, like-for-like comparison of reported performances is often difficult, for example because of differences in cohort recruitment, phenotype definition, population characteristics or use of covariates such as age/sex in reported metrics. Even if the same evaluation cohort was used, many of these interpretation issues remain. To overcome these issues we established a uniform PRS comparison pipeline. This leverages the UK Biobank and several ethnically diverse cohorts (Multi-Ethnic Cohort, BioMe, ARIC, MESA, PAGE) to enable direct comparison between over 50 published and internally developed PRS models for 14 complex diseases in four ancestry groups (European, South Asian, East Asian, African), where sufficient case numbers were available for analysis.
Maximum AUC values were observed for individuals of European ancestries, followed by East Asians (average AUC reduction 2.36%), South Asians (average AUC reduction 4.51%) and Africans (average AUC reduction 10.8%). A few publicly available PRS are optimised for a specific ancestry, but for these we did not observe a predictive boost in the target ancestry. Genome-wide PRS including millions of variants were systematically more effective than sparser models, even when European linkage disequilibrium (LD) was applied to non-European ancestries (e.g. cross ancestry AUC boost of 8.2% in breast cancer). Inclusion of functional annotations yielded slight improvements (average ÎAUC 0.005) across ancestries and diseases. Internal optimization using population specific LD maps and bespoke cross-ancestry methodologies generated a detectable increase in the target ancestries, provided that a training set with an effective sample size â¥20% of the European set was available. For example, type 2 diabetes prediction improved across all ancestries, for a minimum of 10% for African ancestries (ÎAUC 0.06) to a maximum of 16% for East Asian ancestries (ÎAUC 0.09).
Systematic comparisons of PRS models across ancestry groups shows that the expected deterioration of performance was relatively limited between European and East/South Asian ancestries, but much more marked in individuals with African ancestries. Choice of PRS methodology substantially impacted predictive performance, suggesting that methodological developments as well as increasing data availability from non-European studies, will improve PRS performance across ancestries.

PrgmNr 1432 - Power-improved meta-QTL analysis reveals the complex regulation of molecular QTL associated with brain diseases

View session detail

Author Block: **B. Zeng**, J. Bendl, G. E. Hoffman, R. Roussos; Icahn School of Medicine at Mount Sinai, New York, NY
Disclosure Block: **B. Zeng:** None.
Co-localization of risk variants for brain diseases with molecular quantitative trait loci (QTL) aims to refine the credible sets of causal variants and link them to specific genes. Increasing sample size and performing trans-ethnic analysis can increase fine-mapping resolution, and enlarge the power of colocalization analysis. We developed a statistical method, called multivariate multiple QTL (mmQTL), to perform multi-tissue and trans-ethnic large-scale QTL analysis. We first perform dense simulations to demonstrate that mmQTL increases power to detect QTLs, controls the false positive rate in trans-ethnic analysis, and reduces the credible sets of causal variants. We then applied mmQTL on brain bulk RNA-seq (n=3,956 libraries from 2,119 unique donors) and cell-type specific ATAC-seq samples (n=1995 libraries from 646 unique donors) from neuronal and glial cell types. We identify 10,769 eGenes (genes with at least a genome-wide significant eQTL), of which 5,336 eGenes have multiple eQTLs. We detect 17,425 significant caQTL in the neuronal cell, and 15,746 in the glial cell. Functional analysis based on evolutionary scores and rare-variant burden reveals eGenes are more likely to be mutant tolerant than no-eQTL genes, indicating evolutionary constraints on gene expression in the brain. Non-primary eQTLs capture regulatory elements defined by ATAC-seq peaks and are enriched for neuro-degenerative diseases. We found that caQTLs are largely shared among brain regions, but not in cell types. Fine-mapped variants from the power-improved caQTL are enriched to disrupt TF binding sites, and there are different sets of TF genes affected in neuronal and glial cells. Lastly, we integrated the detected eQTL and caQTL to conduct caQTL-eQTL-GWAS colocalization analysis, and revealed the regulation patterns for some disease-associated genes in the human brain. In this study, we have created the largest brain eQTL and caQTL resources to date, and they should be valuable to the community in exploring the underlying biological mechanism in brain diseases.

PrgmNr 1433 - Single-cell RNA sequencing of lung identifies disease-associated cell type-specific eQTLs

View session detail

Author Block: **H. M. Natri**[1], C. B. Azodi[2,3], A. J. Gutierrez[1], L. E. Bui[1], L. Peter[1], R. P. Kendle[1], D. J. McCarthy[2,3], N. E. Banovich[1]; [1]The Translational Genomics Research Institute, Phoenix, AZ, [2]University of Melbourne, Parkville, Victoria, Australia, [3]St. Vincentâ□□s Institute of Medical Research, Fitzroy, Victoria, Australia

Disclosure Block: **H.M. Natri:** None.

Genome-wide association studies and functional genomics approaches have identified regulatory loci that contribute to complex traits and disease. However, the effects of expression quantitative trait loci (eQTL) can be both cell-type and context-specific. Single-cell RNA-sequencing (sc-RNAseq) facilitates the mapping of eQTLs across different cell types, allowing the identification of cell type-specific eQTLs that would go undetected by bulk methods. Here we use sc-RNAseq to examine primary human tissue from the lungs of healthy individuals and those with interstitial lung disease (ILD). ILD is a chronic, progressive lung disease characterized by the scarring of lung tissue through epithelial remodeling and accumulation of extracellular matrix (ECM). Pulmonary Fibrosis (PF) is a clinical phenotype that exhibits the end stage of ILD. In general, progression of fibrosis occurs in a gradient resulting in some regions of extreme remodeling and some regions that appear largely normal. For the most common and severe form of ILD, idiopathic PF (IPF), lung transplantation is the only treatment option.

In the current study, we have used sc-RNAseq to generate expression profiles of 532,488 cells derived from 170 lung tissue samples from 52 healthy and 68 ILD donors, including 39 donors with IPF. To identify cell type-specific expression changes associated with genetic variation, we have used Whole Genome Sequencing (WGS) to genotype 97 individuals with available sc-RNAseq profiles. In an effort to map cis-eQTLs across lung cell types, we have used a pseudo-bulk approach aggregating the expression levels of cells for each of the study individuals. To identify shared and cell type-specific signatures, we have performed stratified analyses as well a joint analysis of all cell types to assess cis-eQTL effect size heterogeneity between cell types. Using this approach, we detect clear cell population, lineage, and cell type-specific signals Further integration with GWAS allows connecting PF risk variants to molecular functions in disease-relevant cell types.

These analyses aid in determining the cell types and states in which PF-associated genetic variants function, and highlight the importance of cellular context in gene regulation in health and disease.

PrgmNr 1434 - The Musculoskeletal 3D Epigenome Atlas

View session detail

Author Block: **M.-J. Tsai**[1], S. Reppe[2], T. Sato[3], R. Gill[4], M. Wein[3], K. Gautvik[2], NHLBI TOPMed Bone Mineralization workgroup, Y.-H. Hsu[1]; [1]Marcus Institute for Aging research and Harvard Medical School, Boston, MA, [2]Oslo University Hospital, Oslo, Norway, [3]Massachusetts General Hospital and Harvard Medical School, Boston, MA, [4]Genentech, San Mateo, CA
Disclosure Block: **M. Tsai:** None.
Musculoskeletal (MSK) disorders are a common and costly problem for elderly populations worldwide, most of which are polygenic disorders. Although an increasing number of human genome-wide association studies (GWAS) and population-based biobank studies identified sequence variants associated with risks of musculoskeletal diseases, there is a lack of information about the cell types in human bone and skeletal muscle tissues. Moreover, it is not a trivial task to understand specific genes regulating the landscape and understanding their functional involvement in musculoskeletal biology, especially for non-coding related variants.To address this issue, we generated HI-C seq, ATAC-seq, and RNA-seq in human primary mesenchymal stem cells, pre-mature osteoblasts, matured osteoblasts, osteocytes, skeletal myoblasts, and myotubes obtained from bone and skeletal muscle biopsies. These data were then combined with publicly available DNase-seq and ChIP-seq in a handful of relevant cell types from the ENCODE and Roadmap Epigenomics projects. We established global maps of regulatory elements and 3D chromatin looping structure with high-resolution (2kb) in human MSK relevant primary cell types. We identified (1) MSK-specific active enhancer-like regions; (2) MSK-specific active transcription factor binding sites and downstream-regulated genes in the open chromatin regions; (3) MSK-specific active promoter-like regions and transcription factors for protein-coding genes; (4) MSK specific proximal and distal enhancer-promoter interactions via the high-resolution 3D loop structure with ATAC-seq and RNA-seq data. The 3D MSK gene regulatory circuits and landscapes are implemented into an online searchable browser, which provides target gene prediction for all available (~700 million) non-coding variants for relevant diseases, visualization for MSK cell-type-specific gene regulatory circuits annotation for every sequence variant observed so far. We also integrated all update-to-date GWAS findings from GWAS Catalog to provide post-GWAS analyses for in-silico functional annotation of associated variants and underlying causal variants and genes prediction. We plan to extend our 3D Epigenome Atlas with additional human primary musculoskeletal cell types with single-cell RNA-seq and single-cell ATAC-seq.In summary, by integrating cell-type-specific multi-omics data, we have established MSK gene regulatory landscapes and developed an online searchable browser that provides comprehensive epigenome annotation and visualization of the 3D genome chromatin interactions. The browser is publicly available.

PrgmNr 1435 - Enhancer enhancer interaction networks involved in long-distance genome regulation link multiple non-coding variants to function

View session detail

Author Block: **X. Lin**; Stanford Univ., Belmont, CA
Disclosure Block: **X. Lin:** None.
Mammalian genomes often use multiple enhancers spanning an ultra long-distance (>Mb) to modulate genes, yet it is not clear how multiple enhancers may coordinate to achieve this task. In addition, Genome-Wide Association Studies (GWAS) reveal that variants in the non-coding regulatory elements, including enhancers, are estimated to account for >90% of the variants in disease. While individual enhancer variants mostly present small-to-modest clinical risks, a combination of multiple genetic variants among various enhancers can greatly amplify the effects of individual low-penetrance variants in complex diseases and traits, suggesting the importance of an enhancer-enhancer interaction network in controlling disease relevant genes and connecting multiple non-coding variants association to function. Recent progress in genomic mapping approaches and CRISPR perturbation technologies have accelerated the identification of enhancer and their roles in gene regulation. However, it remains unclear why multiple enhancers are needed, as well as a quantitative understanding of the underlying enhancer epistasis network, wherein the activity of an enhancer can functionally interact with others.
we used multiplexed CRISPR-based screening with a pooled library consisting of 87,025 sgRNA combinations to combinatorially probe enhancer-enhancer interactions. We discovered a previously uncharacterized two-layer enhancer-enhancer epistasis network across multiple oncogenes. In this network, we defined a class of synergistic regulatory elements (SREs) which are theorized to maintain both expression level and robustness of genes and provide buffering effects against genome instability over long distances. A suite of quantitative experimental and computational approaches, including chromatin interaction assays (Trac-looping), cellular imaging (multicolor 3-dimensional FISH), and machine learning (elastic-net regularized generalized linear model), unveiled mechanisms associated with this enhancer epistasis. We used this information to create a computational model to predict SREs and provided a strategy to link multiple non-coding variants to reveal their epistasis influence in clinical risk. In the genetic association analysis, we found that the SRE variants could cooperatively impact gene expression and alter clinical risk in cancer and autoimmune disorder. Our work unveiled new mechanisms underlying enhancer-mediated control of gene expression in ultra-long genomic distance, with implications for annotation of enhancer function within cells and interpretation of epistasis contribution of non-coding variants in human disease.

PrgmNr 1436 - Escape from X-inactivation in twins exhibits intra- and inter-individual variability across tissues

View session detail

Author Block: **A. Zito**[1,2], A. Visconti[1], N. Rossi[1], A. L. Roberts[1], R. Andres-Ejarque[3], S. Nardone[4], J. E. S. Moustafa[1], M. Falchi[1], K. S. Small[1]; [1]Twin Research & Genetic Epidemiology, King's College London, London, United Kingdom, [2]Present address: Massachusetts General Hospital, Boston, MA, [3]St Johnâ□□s Institute of Dermatology, Kingâ□□s College London, London, United Kingdom, [4]Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA
Disclosure Block: **A. Zito:** None.

X-chromosome inactivation (XCI) silences one X in females to balance the unequal X-linked transcriptional dosage between the sexes. However, over 15% of X-genes escape XCI and is biallelically expressed. Our knowledge of XCI escape in humans largely rely on sex-differences, hybrid cell lines, and epigenetic marks, which are all indirect proxies of escape. Differently, tissue samples with skewed XCI enable the detection and measurement of escape directly in females. At present, the incidence and variability of escape across cells, tissues and individuals are not well known. The extent to which genetics and environment influence escape is also largely unknown. Using RNAseq and DNAseq data in a multi-tissue dataset sampled from 248 skewed female twins, we investigated escape prevalence and variability across fat and skin tissues, lymphoblastoid cell lines (LCLs), and purified immune cells (monocytes, B, T-CD4$^+$, T-CD8$^+$, NK), and individuals.
Solid tissues exhibit up to 13% higher incidence of escape than LCLs. 159 genes escaped XCI in least one tissue. This set includes 54 novel candidate escapees, of which 35% are long-non-coding RNAs. 24 genes constitutively escaped XCI across tissues, while a separate set of 51 genes exhibited tissue-restricted escape. Within an individual, there are both genes escaping XCI in all tissues, and genes showing tissue-restricted escape. Analysis of inter-individual variability revealed 40 genes (e.g. *BTK*, *CD99L2*) exhibiting consistent escape levels across females in at least one tissue, and 62 genes (e.g. *DDX3X*, *KDM6A*, *UBA1*) exhibiting inter-female variability in multiple tissues. Escape is heterogeneous across immune cell types, with higher incidence in lymphocytes than monocytes. There are both genes escaping XCI in multiple immune cell types, and immune cell type-specific escapees. Protein-protein network analysis highlighted that escapees interact with other proteins on a genome-wide scale, and are involved in varied processes and pathways such as epigenetic control of gene activity. Monozygotic (MZ) co-twins share significantly more similar escape levels than dizygotic (DZ) co-twins (corr.MZ=0.6; corr.DZ=0.46; P

PrgmNr 1437 - ChromBPNet: Deep learning models of base-resolution chromatin profiles reveal cis-regulatory syntax and regulatory variation

View session detail

Author Block: **A. Pampari**[1], A. Shcherbina[2], A. Kundaje[3]; [1]Computer Science, Stanford University, Stanford, CA, [2]Biomedical Data Sciences, Stanford University, Stanford, CA, [3]Stanford University, Stanford, CA

Disclosure Block: **A. Pampari:** None.

Chromatin profiling experiments such as DNase-seq, ATAC-seq, and histone ChIP-seq decorate cis-regulatory elements (cREs) with intricate read coverage profiles, whose magnitude, shape and span is regulated by cooperative binding of transcription factors (TFs). Here, we introduce ChromBPNet, the first end-to-end deep learning framework to map DNA sequence to base-resolution chromatin profiles, decipher predictive cis-regulatory sequence syntax of individual cREs and predict the impact of regulatory genetic variants on the strength and shape of different types of chromatin profiles across multiple cell types.

ChromBPNet is an optimized convolutional neural network architecture that models the influence of wide genomic sequence contexts (2-75 kbp) on quantitative base-resolution regulatory profiles from DNase-seq, ATAC-seq and histone ChIP-seq experiments. ChromBPNet trained on five ENCODE canonical cell lines achieved unprecedented predictive performance in held-out chromosomes, while automatically and optimally regressing out assay biases (DNase, Tn5 enzyme bias and input control for ChIP-seq). The models are highly performant over a range of sequencing depths and are able to de-noise and de-sparsify low coverage signal profiles at individual cREs.

We improved interpretation methods for de-novo inference of contribution of individual nucleotides across all putative cREs in the genome, thereby revealing predictive motif instances and their combinatorial interaction effects on base-resolution profiles. We deciphered syntactic sequence heterogeneity of all cREs in each cell-line and benchmarked these predictions against footprinting methods and TF ChIP-seq data. We found remarkable consistency between syntax derived from DNase-seq and ATAC-seq experiments. However, we also found intriguing differences in the influence of motif syntax and their associated cooperative TF complexes on different layers of regulatory activity (TF binding, chromatin accessibility, different types of histone marks) of individual cREs.

Finally, we developed a new variant effect score which predicts the impact of non-coding variants on the strength and shape of base-resolution chromatin profiles, thereby revealing a range of 'blast radii' of variants disrupting different types of TF motifs and syntax. Our models accurately predict quantitative trait loci associated with binding, accessibility and histone marks in lymphoblastoid cell-lines.

Our framework will enable high-resolution annotation of sequence syntax and regulatory variants in 100s of cell types that have been profiled with DNase-seq, ATAC-seq, and histone ChIP-seq experiments.

PrgmNr 1438 - NCBI ALFA Release 2for900 Million Variantsand Allele Frequencyfrom 200KdbGaPSubjects

View session detail

Author Block: **L. Phan**[1], Y. Jin[2], H. Zhang[1], Q. Wang[1], G. Shekhtman[1], D. Shao[1], R. R. Villamarin[1], M. Kimura[1], J. Wang[1], L. Hao[1], N. Sharopova[1], M. Bihan[1], A. Sturcke[1], M. Lee[1], N. Popova[1], W. Wu[1], C. Bastiani[1], M. Ward[3], B. Holmes[1], V. Lyoshin[1], K. Kaur[1], E. Moyer[1], M. Feolo[1], B. L. Kattman[4]; [1]NIH, NLM/NCBI, Bethesda, MD, [2]National Library of Medicine, National Institutes of Health, Bethesda, MD, [3]National Institutes of Health National Center for, Bethesda, MD, [4]NCBI/NLM/NIH, Fort Collins, CO
Disclosure Block: **L. Phan:** None.
NCBI Allele Frequency Aggregator (ALFA) aims to provide the largest and most comprehensive aggregated variant datasets with allele frequency from dbGaP studies as open-access. dbGaP has over two million subjects and up to billions of variants along with thousands of phenotypes and molecular assay datasets. This unprecedented volume and the variety of data hold huge opportunities to exploring and studying genetic variations within human populations and identify genetic factors that influence health and diseases to improve diagnosis, treatment, and prevention.

ALFA Release 2 has over 900 million variants, including 300 million novel variants not in dbSNP Build 154. The data was generated from 79 dbGaP studies that included 192 thousand subjects and 5.8 trillion combined genotypes. Allele frequencies are available for 12 populations, including European, Hispanic, African, Asian, and other diverse population ancestries.

Allele frequencies are available for 86% of variants in dbSNP Build 155 (920M rs), 86% (334K rs) ClinVar small variants, and 99% (22K rs) of variants in the dbGaP GWAS catalog.

This massive amount of data is available as open-access from NCBI for variant interpretation and analysis. It is accessible by web search, FTP download, retrieval using API, and TrackHubs for genomic browsers. Please visit the ALFA homepage for more information about the project, releases, tutorials, and past presentations. ALFA website: https://www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/

PrgmNr 1439 - Phenome-wide association analysis of rare and common variation in 455,000 UK Biobank exome sequences accessible through public data portal

View session detail

Author Block: **Q. Wang**[1], K. Carss[2], R. S. Dhindsa[1], A. Harper[2], A. Nag[2], I. Tachmazidou[2], D. Vitsios[2], S. Deevi[2], J. Okae[2], S. Wasilewski[2], S. Katherine[2], S. Petrovski[2], AstraZeneca Genomics Initiative; [1]Centre for Genomics Research, AstraZeneca, Waltham, MA, [2]Centre for Genomics Research, AstraZeneca, Cambridge, United Kingdom
Disclosure Block: **Q. Wang:** Salary/Employment; AstraZeneca.
We adopted the exome sequences from ~455K UK Biobank participants to study the contribution of both rare and common protein-coding variation to thousands of phenotypic endpoints.
We performed a gene-based collapsing PheWAS and separately a variant-level PheWAS across ~17K binary/quantitative phenotypes, evaluating a range of genetic architectures. These analyses were conducted on our cloud-based platform. Overall, we identified 1,517 gene-phenotype relationships for binary phenotypes and 1,301 gene-phenotype relationships for quantitative phenotypes that were significantly associated (p-9), reflecting 125 and 378 distinct genes, respectively. On comparing with the OMIM database, among the significant findings for binary (clinical) phenotypes, 81% are previously known, 9% are known disease genes but represent novel phenotypic associations, and 10% are novel disease genes. The variant-level analysis yielded 5,178 significant (p-9) non-synonymous variant-phenotype relationships for binary phenotypes and 34,561 significant (p-9) non-synonymous variant-phenotype relationships for quantitative phenotypes, involving 586 and 3,860 distinct genes, respectively. Interestingly, 79% (1,197/1,517) and 46% (596/1,301) of the significant gene-phenotype relationships identified in the collapsing analysis for binary and quantitative phenotypes, respectively, were not detected in the corresponding single-variant analysis, demonstrating the complementarity of the two analytical approaches.
We will illustrate AstraZenecaâ□□s use of these data to rapidly validate and inform safety profiles of potential drug targets, identify novel targets, predict drug repositioning opportunities, and study prevalence of molecular endotypes. We also introduce a data portal to visually navigate this rich resource of PheWAS summary statistics for 19K genes and across 3.2M protein-coding variants. The portal supports the search of associations by gene, phenotype and variant, data visualisations to help interpret and explore the results, and ability to download focused results for further analysis.

PrgmNr 1440 - Integrated multi-omics analysis of thyroid cancer reveals key molecular pathways involved in tumor formation and metastasis

View session detail

Author Block: **S. Wu**[1], A. Sanghi[1], D. Ramazzotti[1], S. Chen[1], L. A. Orloff[1], M. M. Kasowski[1], M. P. Snyder[2]; [1]Stanford University, Palo Alto, CA, [2]Stanford Univ, Stanford, CA
Disclosure Block: **S. Wu:** None.
Thyroid cancer is one of the most common endocrine cancer with a continuously increasing incidence worldwide. Thyroid cancer has the lowest mutational burden among all cancers; however, it exhibits a high degree of heritability. The Cancer Genome Atlas (TCGA) study, one of the most comprehensive studies for thyroid cancer to date, profiled the mutations and transcriptome of 496 primary tumors, and identified two dominant molecular subtypes: BRAF-like and RAS-like. In this study, we performed deep multi-omics profiling (genomics, ATAC-seq, RNA-seq, proteomics, metabolomics, lipidomics) of 27 matched normal thyroid glands, 31 thyroid primary tumors, and 31 lymph node metastasis samples from 36 thyroid cancer patients. We observed dramatic molecular profiling differences between primary tumor/metastasis and normal samples, whereas not between primary tumor and metastasis samples. Transcriptomic, proteomic, metabolomic and lipidomic analyses revealed a dynamic choreography of molecular and cellular events that present three different changing patterns: cancer progression, tumor formation and tumor metastasis. These involve processes such as cell cycle and cell proliferation, cancer-related signaling pathways, immune response, metabolic reprogramming (including Warburg effects, dysregulated energy metabolism, carbohydrates, nucleotides, amino acids and lipids, etc), decreased thyroid hormone biosynthesis. Interestingly, activation of the AhR (aryl hydrocarbon receptor) signaling pathway was found to occur during tumor formation at transcript and protein levels, and we also detected the significant increase of tryptophan catabolite kynurenine, the endogenous tumor-promoting ligand of AhR, from our metabolomics data, indicating AhR signaling pathway and tryptophan catabolism may be the novel drug targets of thyroid cancer. Overall, our results reveal key genomic, molecular and cellular events during thyroid cancer formation and metastasis and its potential molecular mechanisms.

PrgmNr 1441 - Whole Blood RNA Sequencing in a Cohort of Undiagnosed Pediatric Patients

View session detail

Author Block: **H. Hou**[1], L. Kyriakopoulou[2,1], K. E. Yuki[1], S. Barnes[1,3], A. Ramani[4], A. Celik[4], M. Braga[2], M. Gloven-Brown[2], M. Brudno[5], S. S. Meyn[6], G. Costain[7], C. R. Marshall[2,8,9], A. Shlien[1,10], J. Dowling[1,3,11], M. D. Wilson[1,3]; [1]Genetics and Genome Biology, SickKids Research Institute, Toronto, ON, Canada, [2]Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada, [3]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada, [4]The Centre for Computational Medicine, SickKids Research Institute, Toronto, ON, Canada, [5]The Centre for Computational Medicine, The Hospital for Sick Children, Toronto, ON, Canada, [6]University of Wisconsin School of Medicine and Public Health, Madison, WI, [7]The Hospital for Sick Children, Toronto, ON, Canada, [8]The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada, [9]Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, ON, Canada, [10]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada, [11]Department of Paediatrics, University of Toronto, Toronto, ON, Canada

Disclosure Block: **H. Hou:** None.

Background: RNA sequencing (RNA-seq) has the potential to improve our ability to interpret the functional and clinical significance of the genetic variants identified by Whole Exome Sequencing (WES) and Whole Genome Sequencing (WGS). A cohort of 134 pediatric patients with heterogeneous phenotypes collected through the SickKids Genome Clinic (Centre for Genomic Medicine) underwent testing by gene panels, microarray, and subsequently WGS. This resulted in the identification of pathogenic or plausibly causative variants in 44% of the samples. In this study, we set out to assess the utility of blood RNA-seq for the diagnosis of patients with diverse clinical indications and develop an interpretation schema to identify candidate genes.

Methods: We developed a clinical-grade, automated, scalable, and robust end-to-end RNA-seq library preparation platform capable of processing up to 96 samples at one time. We generated whole-blood RNA-seq from 134 individuals using this automated RNA-seq platform. Using a bioinformatics pipeline that incorporates various published tools and custom scripts, we identified expression outliers, aberrant splicing events, and allele-specific expression. We then took a disease gene-centric interpretation approach to identify clinically relevant transcriptional aberrations.

Results: Despite samples being archived for >3 years, we were able to generate good-quality RNA-seq data from all samples. With a median sequencing depth of 115M, we detected >70% of genes in curated genes panels related to phenotypes observed in our cohort. We identified expression outliers and/or aberrant splicing events in clinically relevant genes in 23.1% (31/134) blood RNA-seq data from our patient cohort. Specifically, we found RNA evidence supporting previously identified genetic variants in 30.5% (18/56) patients and proposed candidate genes in 16.6% (13/78) cases with negative WES/WGS findings. For example, we identified a known Joubert syndrome gene CEP120 as a decreasing expression outlier that contained a deep intronic splicing alteration in a patient with global developmental delay, abnormality of brain morphology, and dysmorphic facial features.

Conclusion: Our results support the use of clinical blood RNA-seq to facilitate genome diagnostics in pediatric patients with diverse phenotypes.

PrgmNr 1442 - Long read sequencing reveals extensive structural variation in diverse mouse genomes

View session detail

Author Block: **A. Ferraj**[1,2], P. A. Audano[2], P. Balachandran[2], I. A. Walawalkar[2], A. Czechanski[3], L. G. Reinholdt[3], C. R. Beck[1,2]; [1]University of Connecticut Health Center, Farmington, CT, [2]The Jackson Laboratory for Genomic Medicine, Farmington, CT, [3]The Jackson Laboratory, Bar Harbor, ME
Disclosure Block: **A. Ferraj:** None.
Structural variants (SVs) are genomic rearrangements that are larger than 50 bp in length. SVs consist of insertions, deletions, duplications, inversions, and complex combinations of these events; they collectively have implications in the pathology of cancer, Mendelian diseases, phenotypic variation, and evolution. Whole genome discovery and characterization of SVs has primarily utilized Illumina short-read sequencing, which can lack sensitivity in GC-rich regions and repetitive regions and can falter in the detection of some SVs, including inversions, complex SVs, and insertions. With mobile element polymorphisms, short reads cannot resolve the internal structure and sequence of these events when they are identified. Short reads have been used to study SVs in thousands of human and mouse genomes. In recent years, it has been shown that long-read sequencing excels at identifying many SVs, particularly insertions and duplications; these technologies can also resolve the nucleotide sequence of longer insertions. While the Human Genome Structural Variation Consortium (HGSVC) has released the most complete SV call sets to date, similar long-read efforts have not been applied to mouse strains, which leaves a gap in our understanding of a model organism critical to basic and clinical research. With Pacific Biosciences (PacBio) long-read sequencing, we are annotating and characterizing SVs in 13 widely-utilized inbred laboratory mouse strains. With an ensemble of long read SV calling algorithms, we have identified 495,452 SVs which occur at unique sites across the current mouse reference assembly, cumulatively affecting 10% (300Mb) of the mouse genome. We find that SVs account for over 5x the number of nucleotide changes between strains when compared to single nucleotide variants. Most of novel SVs found (90%) were specific to long-read detection. Furthermore, we have detected strain-specific as well as inter-strain SVs that exhibit high impact consequences, including over 2,300 coding sequence variants, 300 frameshift events, and insertion variants that dramatically affect the transcripts that they occur within. Our data comprehensively identify the extensive SVs present in diverse mouse genomes and strongly suggest that SVs contribute to inbred mouse genomic and phenotypic diversity.

PrgmNr 1443 - MERSCOPE™ reveals the transcriptional organization of the mouse brain

View session detail

Author Block: **G. Emanuel**, J. He, J. L. He; Vizgen, Cambridge, MA
Disclosure Block: **G. Emanuel:** Salary/Employment; Vizgen.
Biological systems are comprised of numerous cell types, intricately organized to form functional tissues and organs. Building molecular atlases to fully understand the structure and function of each cell within the brain is now a key aspect of neuroscience research. Atlas initiatives using single-cell RNA sequencing can characterize cell types based on their RNA expression profiles. However, the tissue organization is lost when cells are dissociated for single-cell sequencing, making it difficult to study how the cellular heterogeneity is contributing to the function of the tissue. Furthermore, accurately characterizing each cell within the brain is challenging due to the low expression of many functionally important genes such as nonsensory G-protein coupled receptors (GPCRs). A true spatial transcriptomics technology with high detection efficiency and single-molecule resolution is required to build accurate and complete molecular atlases. Vizgenâ□□s in situ genomics platform MERSCOPE™ enables the direct profiling of the spatial organization of intact tissue with subcellular resolution. MERSCOPE is built on multiplexed error robust in situ hybridization (MERFISH) technology that uses combinatorial labeling, sequential imaging, and error-robust barcoding to provide the highest detection efficiency and resolution available for spatial genomics. In a single experiment, hundreds of thousands of cells can be spatially profiled with high accuracy and reproducibility. To demonstrate the power of MERSCOPE, we mapped 483 genes across three full mouse brain coronal slices. We constructed a panel of canonical cell type markers and nonsensory GPCRs to spatially profile nonsensory GPCR expression across the brain with cellular context. Nonsensory GPCRs in the brain mediate signaling and may play vital roles behind brain ageing and neurodegenerative disorders but are difficult to analyze. Our experiment successfully detected multiple lowly expressed GPCRs including Oxtr, Tshr, and Insr. The mouse brain receptor map demonstrates MERSCOPE as a leading tool for molecular atlassing, enabling scientists to find greater insights into healthy versus diseased tissue.

PrgmNr 1444 - Identifying the genetic contributors of efficacy and adverse metabolic effects of thiazide diuretics in African Americans from the Genetics of Hypertension Associated Treatments (GenHAT) study

View session detail

Author Block: **N. M. Davis Armstrong**[1], V. Srinivasasainagendra[1], R. M. Tanner[1], B. A. Hidalgo[1], H. K. Tiwari[1], D. K. Arnett[2], M. R. Irvin[1]; [1]University of Alabama at Birmingham, Birmingham, AL, [2]University of Kentucky, Lexington, KY

Disclosure Block: **N.M. Davis Armstrong:** None.

African Americans (AAs) have a higher prevalence of hypertension (HTN) compared to whites, presenting with a more severe form due to earlier onset and more rapid vascular damage. AAs respond better to diuretics compared to beta blockers or ACE inhibitors, yet the reason is not well understood. While thiazide diuretics continue to be a first-line antihypertensive, there is clinical significance to adverse metabolic effects linked to thiazide use, including incident diabetes and hypokalemia. The GenHAT study is an ancillary study of the Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). DNA from participants was extracted and hybridized to Illumina Multi-Ethnic AMR/AFR BeadChip arrays. Quality control was performed at the sample and variant level, resulting in the inclusion of 4,297 AAs taking chlorthalidone, a thiazide diuretic, and 969,031 genotyped variants. Upon imputation using the NHLBI TOPMed Freeze 8 reference panel, over 20 million variants with minor allele counts >20, imputation quality scores >0.3, and genotype probabilities >90% were retained for association analysis. Outcomes of interest included systolic (SBP) and diastolic (DBP) blood pressure response over 6 months, fasting glucose response (FG) over 24 months, and serum potassium response (K) over 2 months. Linear regression models for the response of all outcomes were performed in PLINK2 and adjusted for age, sex, baseline measure, and genetic ancestry. A total of 14 variants (3 SBP, 2 DBP, and 9 FG) exceeded statistical significance at pCDHR2 and *MINDY3-CUBN* gene regions for SBP, *LINC02211-CDH9* intergenic region for DBP, and *GIMAP1-GIMAP5* for FG analyses. While many of the variants identified had no previous association with HTN or drug response, variants located within *CDH9*, *MINDY3-CUBN*, and *GIMAP* gene regions were of particular interest. *CDH9* variants have been previously associated with coronary artery calcification in an AA meta-analysis, while variants in *CUBN* have been linked to albuminuria and coronary artery disease. The *GIMAP* GTPase family are potential modifiers in autoimmune diseases including diabetes, and have reported associations with triglycerides, von Willebrand factor, C-reactive protein, and fibrinogen levels. Replication for these variants is ongoing using data from the International Consortium for Antihypertensive Pharmacogenomics Studies (ICAPS) to determine whether these variants can help identify individuals at risk of poorer drug response or unfavorable metabolic changes.

PrgmNr 1445 - Paradigm Shift from Disease PRS to PGx PRS for Drug Response Prediction using PRS-PGx Methods

View session detail

Author Block: **S. Zhai**, H. Zhang, J. Shen; Biostatistics and Research Decision Sciences, Merck Research Laboratories, Merck & Co., Inc., Rahway, NJ
Disclosure Block: **S. Zhai:** None.
Polygenic risk score (PRS), by combining many small prognostic genetic effects, has shown promise in predicting human diseases and complex traits. In efficacy-based pharmacogenomics (PGx) studies, the current practice is to apply such PRS built from disease GWAS directly to PGx data from randomized clinical trials for drug response prediction and patient stratification. However, this approach relies on the assumption that every variant selected for constructing PRS has a constant ratio between its genotype main and genotype-by-treatment interaction effects, which largely may not be true in real PGx data. A violation of such assumption will make disease PRS explain less heritability of drug responses and thus reduce power in predicting them. Here, we propose the paradigm shift from disease PRS to PGx PRS approaches by simultaneously modeling the prognostic and predictive effects and constructing both PRSs for drug response prediction in PGx. We make this paradigm shift possible by developing a series of novel PRS-PGx methods, including PRS-PGx-unadj (unadjusted), PRS-PGx-CT (Clumping + Thresholding), PRS-PGx-L, -GL, -SGL (Lasso-, Group Lasso-, Sparse Group Lasso-based penalized regression), and PRS-PGx-Bayes (Bayesian regression). In the framework of Bayesian regression, we propose a polygenic prediction method that infers posterior prognostic and predictive effect sizes of SNPs simultaneously using PGx genome-wide association summary statistics and an external linkage disequilibrium (LD) reference panel. By introducing global-local continuous shrinkage priors on SNP effect sizes, our proposed PRS-PGx-Bayes method is more robust to varying relationships between the genotype main and genotype-by-treatment interaction effects. Extensive simulation studies show that PRS-PGx methods generally outperform the current disease PRS (PRS-Dis) methods across a wide range of genetic architectures and PRS-PGx-Bayes is superior to all other PRS-PGx methods. We further apply the PRS-PGx methods to the IMPROVE-IT PGx GWAS data by constructing PGx PRSs via 5-fold cross-validation to predict low-density lipoprotein-cholesterol. The drug response prediction results demonstrate the great improvement of PRS-PGx-Bayes in both prediction accuracy and the capability of capturing the treatment-specific predictive effects over alternative methods.

PrgmNr 1446 - Utility of polygenic risk scores for colorectal cancer risk assessment across diverse populations

View session detail

Author Block: **M. Thomas**[1], L. C. Sakoda[2], M. Timofeeva[3], V. DÃez-Obrero[4], Z. Chen[5], P. Law[6], E. A. Rosenthal[7], S. Schmit[8], C. Fernandez-Rozadilla[9], C. R. Gignoux[10], J. C. Figueiredo[11], J. Huyghe[12], T. A. Harrison[12], R. B. Hayes[13], I. Tomlinson[14], S. B. Gruber[11], G. Casey[15], R. S. Houlston[16], V. Moreno[17], W. Zheng[18], M. Dunlop[19], G. P. Jarvik[20], D. A. Corley[21], U. Peters[22], L. Hsu[12]; [1]Fred Hutch, Seattle, WA, [2]Kaiser Permanente, Oakland, CA, [3]University of Southern Denmark, Odense, Denmark, [4]Colorectal Cancer Group, Molecular Mechanisms and Experimental Therapy in Oncology (ONCOBELL) Program, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain, [5]Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN, [6]Division of Genetics and Epidemiology, The Institute for Cancer Research, London, United Kingdom, [7]University of Washington School of Medicine, Seattle, WA, [8]Cleveland Clinic, Cleveland, OH, [9]Instituto de Investigacion Sanitaria de Santiago - IDIS, Santiago de Compostela, Spain, [10]Univ of Colorado Denver, Anschutz Medical Campus, Aurora, CO, [11]USC, Los Angeles, CA, [12]Fred Hutchinson Cancer Research Center, Seattle, WA, [13]Division of Epidemiology, Department of Population Health, New York University School of Medicine, New York, NY, [14]Cancer Research UK Edinburgh Centre, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom, [15]Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, VA, [16]The Institute of Cancer Research, Sutton, United Kingdom, [17]Oncology Data Analytics Program, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain, [18]Vanderbilt Univ Med Center, Nashville, TN, [19]Colon Cancer Genetics Group, Medical Research Council Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, United Kingdom, [20]Univ Washington Med Center, Seattle, WA, [21]Division of Research, Kaiser Permanente Northern California, Oakland, CA, [22]Fred Hutchinson CA Res Ctr, Seattle, WA
Disclosure Block: **M. Thomas:** None.

Colorectal cancer (CRC) is a leading cause of cancer-related mortality, yet many CRC are preventable by screening individuals at increased disease risk. Polygenic risk score (PRS) offer the prospect of precision prevention by defining individualâ□□s risk of CRC and tailoring screening. Genome-wide CRC-PRS built upon European (EUR) data (PMID: 32758450) have limited performance in non-EUR populations because of small number of non-EUR populations included in the training data set. To address this deficiency, we expanded our PRS development to include non-EUR populations.
We derived PRS by leveraging GWAS summary statistics and LD structure from EUR and East Asian (EAS) ancestry. Our training data contains 78,473 cases & 107,142 controls of EUR ancestry and 21,737 cases & 47,444 controls of EAS ancestry. We examined 3 approaches for PRS development, using 1) 180 CRC known loci, 2) genome-wide EAS and EUR summary statistics and LD matrices, 3) combined genome-wide EUR and EAS summary statistics and weighted LD matrices with weights defined as the proportion of subjects from each ancestry. For Approaches 2 and 3, we used LDpred2 to derive PRS including ~1M SNPs. For Approaches 1 and 3 we derived a single PRS across EAS and EUR populations, whereas for Approach 2 we first derived ancestry-specific PRSs (PRSEUR and PRSEAS) and then final PRS for each ancestry by leveraging PRS from other ancestry using a weighted sum of PRSEUR and PRSEAS where the weights were obtained from a multivariate regression model, using 2,627 cases & 3,797 controls for EAS and 29,864 cases & 31,629 controls for EUR ancestry. To evaluate the performance, we analyzed independent data from Genetic Epidemiology Research on Adult Health and Aging cohort of 101,987 individuals, which included 1,699 CRC cases.
The AUC of the best performing approach, which combined genome-wide EUR and EAS summary statistics and LD matrices (Approach 3), were 0.67, 0.65, 0.63 and 0.57 for European, East Asian, Latinx, and African American ancestry groups, respectively, with hazard ratios of top 30% vs.

remaining for corresponding groups of 2.52, 1.61, 2.15 and 1.37. Compared to EUR-based PRS (PMID: 32758450), the AUCs for Approach 3 are improved by 3%, 6%, 5% and 2% for European, East Asian, Latinx and African American ancestry groups, respectively.
A trans-ethnic PRS has the potential to reduce disparities in PRS performance between non-EUR and EUR which is critical if PRS is going to be implemented in clinical practice or populations screening programs. As the performance remains lower in non-EUR populations, larger and more racially and ethnically diverse study populations may further improve equitable risk prediction.

PrgmNr 1447 - Dissecting 90 percent of lipoprotein(a) heritability in 487,571 UK Biobank participants

View session detail

Author Block: **R. E. Mukamel**[1], R. E. Handsaker[2], M. A. Sherman[3], A. Barton[4], Y. Zheng[1], S. A. McCarroll[4], P.-R. Loh[5]; [1]Brigham and Women's Hospital/Harvard Medical School, Boston, MA, [2]Charlemont, MA, [3]Massachusetts Institute of Technology, Marblehead, MA, [4]Harvard Medical School, Boston, MA, [5]Brigham and Women's Hospital / Harvard Medical School, Boston, MA
Disclosure Block: **R.E. Mukamel:** None.
Lipoprotein(a) concentration [Lp(a)] is a highly heritable (h^2>90%), monogenic cardiovascular risk factor whose primary genetic determinant is the size of the *LPA* kringle IV-2 (KIV-2) repeat, which explains ~60% of Lp(a) heritability. However, the full spectrum of other functional sequence variants in *LPA* and the way in which they shape Lp(a) variation has been unclear due to the difficulty of genotyping KIV-2 variation from high-throughput genomic assays. We developed methods to accurately estimate haplotype-resolved KIV-2 allele lengths in N=49,959 exome-sequenced UK Biobank (UKB) participants (RMSE ~1 repeat unit), and to impute KIV-2 lengths into SNP haplotypes for the remaining N=437,612 UKB participants (imputation R^2=0.9).
To systematically identify and measure the effects of additional Lp(a)-altering variants, we isolated the contributions of individual *LPA* haplotypes by analyzing heterozygous carriers of two coding variants known to create nonfunctional alleles (rs41272114 and rs41259144, combined MAF=0.05). This approach created an effective haploid model for Lp(a). Stepwise conditional analyses controlling for KIV-2 length pointed to a series of 20 additional protein-altering and 5â UTR variants that appeared to further shape Lp(a). Together with KIV-2 length, these variants explained 90% of heritable Lp(a) variance in European-ancestry participants in a model that accounted for nonlinear and cis-epistatic effects. These variants also largely explained the ~4-fold variation in median Lp(a) across populations: the higher Lp(a) commonly observed in African populations appeared to result from a relative paucity of alleles carrying a large-effect Lp(a)-reducing coding or splice-altering SNP in LPA (affecting 13% vs. 43% for African vs. European alleles) and the high frequency of the Lp(a)-increasing 5â UTR variant rs1800769 (46% vs. 17% for African vs. European alleles).
The accuracy of genetically predicted Lp(a) also enabled insights into epidemiological associations with Lp(a). The myocardial infarction risk-increasing effect of higher Lp(a) extended to extreme Lp(a) levels in UKB (OR=3.1, 95% CI=1.9-5.2 for individuals with predicted Lp(a)>400 nmol/L). In contrast, lower genetically predicted Lp(a) did not associate with increased type-2 diabetes (T2D) risk, suggesting that the 17% (s.e. 1%) lower levels of measured Lp(a) observed in T2D patients represents reverse causation resulting from T2D itself, T2D-related liver comorbidities, or T2D medication.
These results demonstrate that complex genetics can underlie even monogenic traits and inform genetic approaches to cardiovascular risk stratification.

PrgmNr 1448 - Polygenic score from a large GWAS predicts cases of heart failure with reduced ejection fraction (HFrEF) but not preserved ejection fraction (HFpEF)

View session detail

Author Block: **K.-H. Wu**[1], N. J. Douville[2], B. N. Wolford[1], I. Surakka[1], W. E. Hornsby[1], M. C. Konerman[2], M. R. Mathis[2], M. J. Daly[3], B. M. Neale[3], C. J. Willer[1], W. Zhou[4], the Global Biobank Meta-analysis Initiative; [1]University of Michigan, Ann Arbor, MI, [2]Michigan Medicine, Ann Arbor, MI, [3]Massachusetts General Hospital, Boston, MA, [4]Massachusetts general hospital, Broad Institute, Boston, MA
Disclosure Block: **K. Wu:** None.
Introduction Heart failure (HF) is a complex disease with multiple subtypes that have distinct etiologies, pathophysiologies, and genetic risk. The performance of a heart failure polygenic score (PGS), derived from a new genome-wide association study (GWAS), was tested to predict cases of heart failure with reduced ejection fraction (HFrEF) and preserved ejection fraction (HFpEF). **Methods** GWAS was performed within the Global Biobank Meta-analysis Initiative (GBMI) â a global collaboration among 13 biobanks across the world with diverse ancestries. GWAS summary statistics were then used to generate a PGS in a combined cohort from the Michigan Genomics Initiative and Cardiovascular Health Improvement Project (MGI/CHIP). Heart failure cases in the GBMI training dataset were defined based upon ICD codes, which did not distinguish between HF subtypes. Electronic health record data available within MGI/CHIP enabled further classification of patients into HFrEF and HFpEF, using a previously validated methodology incorporating ICD diagnoses, free-text language processing, and left ventricular ejection fraction (LVEF). Twenty percent of the cases derived with this method were further adjudicated by clinicians to confirm the phenotype. To compare the predictive ability of PGS, we evaluated logistic regression models with PGS adjusted for age, sex, and principal components derived from genotype data separately for both HFrEF and HFpEF phenotypes. **Results** In the GBMI training dataset, genetic data was analyzed from a total of 67,049 HF patients from 1,305,592 samples from 6 ancestral populations: 25.4% of the samples were of non-European ancestry. The GWAS identified 22 index variants that reached genome-wide significance. The MGI/CHIP validation dataset contained: 360 HFrEF patients, 232 HFpEF patients, and 24,313 healthy controls. The genome-wide PGS is a significantly better predictor of HFrEF compared to HFpEF in the MGI/CHIP cohort. In the HFrEF model, the PGS had an adjusted odds ratio (aOR) of 1.40 (95% CI: 1.25-1.57; p-value: $1.25x10^{-8}$) compared to an aOR of 1.08 (95% CI: 0.93-1.24; p-value: 0.30) in the HFpEF model. **Conclusion** Our analyses showed that a PGS for heart failure derived from GBMI data is useful in predicting HFrEF in an independent dataset. The difficulty in predicting HFpEF could result from: (i) the GBMI HF phenotype preferencing HFrEF over HFpEF, (ii) increased diagnostic accuracy in HFrEF in the evaluation cohort (due to inclusion of decreased LVEF), or (iii) greater genetic heterogeneity in the HFpEF population. Future studies focused on a GWAS for HFpEF may create a more useful polygenic score, if the trait is sufficiently heritable.

PrgmNr 1449 - Rare variant carrier status on risk of coronary artery disease among individuals with high polygenic risk

View session detail

Author Block: **M. Shivakumar**[1], S.-H. JUNG[2], J.-S. Yun[3], M. D. Ritchie[2,4], D. Kim[2]; [1]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA, Philadelphia, PA, [2]University of Pennsylvania, Philadelphia, PA, [3]The Catholic University of Korea, Bryn Mawr, PA, [4]Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA
Disclosure Block: **M. Shivakumar:** None.

Polygenic risk scores (PRS) derived from large GWAS studies have been able to capture a large portion of genetic risk for coronary artery disease (CAD) risk. Previous studies have shown that the PRS for CAD (PRS_{CAD}) can be used to identify the subgroup from the population that is at greater than three-fold increased risk for CAD. However, in general, rare variants are known to have larger impacts than common variants, and complex diseases like CAD are influenced by both common and rare genetic variants. In this study, we examined the changes in the risk conferred by carriers of rare variants in known genes associated with CAD.

Among 154,799 unrelated individuals of European ancestry, we identified 2,553 CAD cases with both genotype and exome sequencing data available in UK Biobank (UKBB). PRS_{CAD} for all samples were calculated using genotype data using LDpred and summary statistics from the CARDIoGRAMplusC4D GWAS study. The rare variants were obtained from the exome sequencing data and were filtered to keep only predicted loss of function (LOF) variants, variants predicated with high impact by Variant Effect Predictor, and missense variants with REVEL score > 0.5. We split samples into 10 groups based on 5 PRS_{CAD} quantiles and rare variant carrier status. We calculated the risk of CAD between the first and last quantile combined with rare variant carrier status.

Of the 47 known replicated genes for CAD, we found 22 genes where the carriers of rare variants showed higher odds of CAD risk than non-carriers, and in the other 25 genes, the carriers of rare variants showed lower odds of CAD risk than the non-carriers in the high risk PRS group. The samples with LOF rare variants in genes such as *MIA3* (Lowest vs. highest PRS with non-carrier OR: 2.00, carrier OR: 4.13) and *TCF21* (non-carrier OR: 2.00, carrier OR: 3.09) showed higher OR. The high expression of MIA3 is expected to promote atheroprotective vascular smooth muscle cells and higher expression *TCF21* is known to inhibit risk for coronary artery disease. On the contrary, genes like *SMAD3* (non-carrier OR: 2.02, carrier OR: 1.25E-05) and HHIPL1 (non-carrier OR: 2.01, carrier OR: 1.29E-05) showed lower odds, and their expression is known to increase CAD progression.

We showed that the LOF rare variant carriers in the high risk PRS groups may increase or decrease the odds of CAD risk. This study also shows the need to take rare variants into consideration for risk stratification. Robust methods for integration of rare variants with PRS are greatly needed to further improve risk prediction for CAD and allow for the development of improved preventive measures and also novel therapeutic opportunities for CAD.

PrgmNr 1450 - Expanded Genetic Clustering of Type 2 Diabetes Loci Using a High-throughput Pipeline Reveals Mechanistic Pathways of Metabolic Diseases

View session detail

Author Block: **H. Kim**[1,2], K. Westerman[1,2], J. Chiou[3], J. B. Cole[2,1,4,5], T. D. Majarian[2], M. von Grotthuss[6], J. M. Mercader[2,1,4], S. Kwak[7], J. Kim[8], J. C. Florez[1,2,4], K. Gaulton[9], A. K. Manning[1,2,4], M. S. Udler[10,1,4]; [1]Massachusetts General Hospital, Boston, MA, [2]The Broad Institute of MIT and Harvard, Cambridge, MA, [3]UC San Diego, La Jolla, CA, [4]Harvard Medical School, Boston, MA, [5]Boston Children's Hospital, Boston, MA, [6]Takeda Pharmaceuticals, Cambridge, MA, [7]Seoul National University Hospital, Seoul, Korea, Republic of, [8]GlaxoSmithKline, Cambridge, MA, [9]UC San Diego, San Diego, CA, [10]Broad Institute, Cambridge, MA

Disclosure Block: **H. Kim:** None.

Complex diseases such as type 2 diabetes (T2D) are highly polygenic and influenced by multiple biological pathways. Rapid expansion in the number of T2D loci can be leveraged to identify such pathways, which may facilitate improved patient management.

We developed a high-throughput pipeline to enable clustering of T2D loci based on variant-trait associations. Our pipeline extracted summary statistics from genome-wide association studies (GWAS) for T2D and related trait to generate a matrix of 336 variants x 64 trait associations and applied Bayesian Non-negative Factorization (bNMF) to identify genetic components of T2D.

We identified ten genetic clusters of T2D, which included five from our published prior analysis of 94 T2D loci. Four of the ten clusters related to mechanisms of insulin deficiency, five to insulin resistance, and one had an unclear mechanism. Novel clusters identified in this analysis related to beta-cell dysfunction, pronounced insulin secretion, and circulating levels of alkaline phosphatase, lipoprotein A, and sex hormone binding globulin.

The T2D genetic clusters displayed tissue-specific epigenomic enrichment, particularly in pancreatic islets, liver, and adipose tissue. Two of the clusters relating to insulin deficiency were differentially enriched in functional and stressed pancreatic beta-cell states. Additionally, cluster-specific polygenic scores were associated with distinct clinical outcomes across GWAS and confirmed in participants in the Mass General Brigham Biobank. Multiple observed T2D genetic pathways were shared across genetic clusters of coronary artery disease and chronic kidney disease.

Our approach stratifies T2D loci into physiologically meaningful clusters with distinct tissue specificity and association with metabolic conditions. The pipeline allows for efficient updating and refining of clusters as additional GWAS datasets become available, and can be readily applied to other conditions. This method supports translation of GWAS findings into a more granular understanding of disease mechanisms, with a view toward precision medicine.

PrgmNr 1451 - Mirror effects of *OPRD1* variants on diabetes and obesity

View session detail

Author Block: **S. Meulebrouck**[1,2], G. QuÃ©niat[1,2], M. Baron[1,2], M. Canouil[1,2], B. Balkau[3], G. Charpentier[4], S. Franc[4,5], M. Marre[6], R. Roussel[6,7,8], R. Scharfmann[9], A. Bonnefond[1,2,10], P. Froguel[1,2,10]; [1]Inserm UMR1283, CNRS UMR8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, Lille University Hospital, Lille, France, [2]UniversitÃ© de Lille, Lille, France, [3]Paris-Saclay University, Paris-Sud University, UVSQ, Center for Research in Epidemiology and Population Health, Inserm U1018, Clinical Epidemiology, Villejuif, France, [4]CERITD (Centre dâ□□Ã□tude et de Recherche pour lâ□□Intensification du Traitement du DiabÃ¨te), Evry, France, [5]Department of Diabetes, Sud-Francilien Hospital, University Paris-Sud, Corbeil-Essonnes, France, [6]Inserm U1138, Centre de Recherche des Cordeliers, Paris, France, [7]Department of Diabetology Endocrinology Nutrition, HÃ´pital Bichat, DHU FIRE, Assistance Publique HÃ´pitaux de Paris, Paris, France, [8]UFR de MÃ©decine, University Paris Diderot, Sorbonne Paris CitÃ©, Paris, France, [9]Institut Cochin, INSERM, U1016, CNRS, UMR8104, UniversitÃ© de Paris, Paris, France, [10]Department of Metabolism, Imperial College London, London, United Kingdom

Disclosure Block: **S. Meulebrouck:** None.

Opioid consumption leads to contradictory effects on metabolic homeostasis, by increasing hyperglycemia but improving lipid profile and adiposity. However, the mechanisms linking opioids and metabolism are unknown. RNA sequencing data showed that *OPRD1* encoding Î´ opioid receptor (DOP) is expressed in human islets, especially in Ã□ cells. DOP is an inhibitory G-protein coupled receptor. Based on *OPRD1* resequencing and functional experiments, we aimed to decipher the putative link between *OPRD1* mutations and metabolic disorders.

OPRD1 was sequenced in 6,971 individuals. The effect of each detected variant was assessed via luciferase experiments, in response to increasing concentrations of two DOP agonists (DII or DPDPE). We categorized these variants as gain-of-function (GoF) or loss-of-function (LoF). In parallel, expression and localization of each variant were assessed by western blotting and immunofluorescence assays. Association studies between GoF or LoF variants and various metabolic traits were assessed in our cohort, and in further 34,812 individuals from the T2D Knowledge Portal for the study of the frequent GoF variant encoding p.I52V. Finally, we performed glucose-stimulated insulin secretion (GSIS) in the human Ã□-cell model EndoCÃ□H1 overexpressing *OPRD1* and treated with DII DOP agonist.

In 6,971 individuals, we detected 31 rare variants and 3 frequent variants of *OPRD1*. Luciferase assays highlighted 7 GoF variants, including the frequent variant encoding p.I52V, and 12 LoF variants. Immunofluorescence assays showed that all the mutants were effectively expressed and localized at the plasma membrane, except for two LoF mutants (p.P14R and p.G36E). Western blots showed that these two mutants tended to have a lower expression than wild-type DOP. Association analyses revealed that rare LoF variants increased overweight and obesity risk (P=0.0054; OR=11) but decreased hyperglycemia risk (P=0.054; OR=0.23), while rare GoF variants significantly improved lipid profile. Besides, the frequent GoF p.I52V variant was associated with increased type 2 diabetes risk ($P=3.6 \times 10^{-6}$; OR=2) but decreased body mass index (P=0.0038; Ã□=-0.37) and improved lipid profile, confirming this mirror effect. Finally, we showed that DOP significantly inhibited insulin secretion from Ã□ cells.

This study highlights DOP as a major link between opioids and metabolic disorders. DOP agonists and/or antagonists should be considered as new tools to improve metabolic homeostasis.

PrgmNr 1452 - The effect of obesity-related traits on COVID-19 severe respiratory symptoms and hospitalization and its mediation by socioeconomic status: a multivariable Mendelian randomization study

View session detail

Author Block: **B. Cabrera Mendoza**[1,2], F. Wendt[1,2], G. Pathak[2,1], F. de Angelis[1,2], A. De Lillo[1], D. Koller[1,2], R. Polimanti[1,2]; [1]Yale University, New Haven, CT, [2]VA CT Health Care System, West Haven, CT
Disclosure Block: **B. Cabrera Mendoza:** None.
Obesity has been associated with a higher susceptibility to coronavirus disease 2019 (COVID-19), particularly with its more severe clinical manifestations. However, this association can be affected by many correlates of these traits. Due to its large impact on human health, socioeconomic status (SES) could influence at least partially the association between obesity and COVID-19 severity. To estimate the independent effect of traits related to body size and SES on the clinical manifestations of COVID-19, we conducted a Mendelian randomization (MR) study analyzing the effect of obesity-related anthropometric traits on COVID-19 outcomes. We evaluated the effects of body mass index (BMI), waist circumference (WC), hip circumference, (HIP) and waist-hip ratio (WHR) studied in up to 234,069 participants from the Genetic Investigation of ANthropometric Traits (GIANT) consortium with respect to three COVID-19 outcomes: severe respiratory COVID-19 (5,101 cases vs. 1,383,241 controls), hospitalized COVID-19 (9,986 cases vs. 1,877,672 controls), and COVID-19 infection (38,984 cases vs. 1,644,784 controls) obtained from the COVID-19 Host Genetics Initiative (HGI). Finally, to investigate the effect of SES, we analyzed genetic data related to self-reported household income (HI) information from 286,301 UK Biobank (UKB) participants. We found that BMI and WC were associated with severe respiratory COVID-19 (BMI: ORâ□□=â□□1.68 pâ□□=â□□0.0004; WC: ORâ□□=â□□1.72, pâ□□=â□□0.0072) and hospitalized COVID-19 (BMI: OR =â□□1.62, pâ□□=â□□1.35e-06; WC: ORâ□□=â□□1.62, pâ□□=â□□0.0001). Also, HIP showed to influence hospitalized COVID-19 (ORâ□□=â□□1.31, pâ□□=â□□0.012) and COVID-19 infection (ORâ□□=â□□1.18, pâ□□=â□□0.0016). Conversely, HI was associated with reduced severe respiratory COVID-19 (ORâ□□=â□□0.57, pâ□□=â□□0.011) and hospitalized COVID-19 (ORâ□□=â□□0.71, pâ□□=â□□0.045). Testing these effects in multivariable MR models, we observed that the effect of obesity-related anthropometric traits on COVID-19 outcomes is not independent of the SES effect assessed as HI.. In summary, our findings indicate that low SES is a contributor to the observed association between body size and COVID-19 outcomes. Thus, the association of obesity with COVID-19 outcomes may be due to the conditions related to low SES rather than pathogenic mechanisms linked to obesity. This result has major public health implications because it supports that preventive strategies targeting body size and composition to reduce COVID-19 morbidity and mortality may not be effective if they are not considered in the context of SES.

PrgmNr 1453 - Comparative genetic architecture of the inflammatory bowel disease across East Asian and European populations

View session detail

Author Block: **R. Liu**[1,2], H. Gao[3], M. Li[4], H. Huang[2,1], Z. Liu[3]; [1]Broad Institute of MIT and Harvard, Cambridge, MA, [2]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, [3]Center for IBD Research, Department of Gastroenterology, Shanghai Tenth Peopleâ□□s Hospital of Tongji University, Shanghai, China, [4]Department of Gastroenterology, Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, China
Disclosure Block: **R. Liu:** None.
The inflammatory bowel diseases (IBD) are a group of chronic autoimmune disorders including two major subtypes: Crohnâ□□s disease and ulcerative colitis. The prevalence and incidence of IBD are increasing worldwide, especially in East Asia and other developing countries. Genome-wide association studies have identified over 240 of risk loci for IBD. However, the majority of IBD genetic studies were conducted using subjects of European descent (EUR), limiting the discovery and application of IBD genetics to the rest of the world populations. To address this issue, we conducted a large-scale IBD genetic study in the East Asian populations (EAS) using subjects from China, Japan and Korea, for a total sample size close to 18,000 (CD: 4,323, UC: 3,648, and Control: 10,014). All subjects were genotyped on the Illumina Immunochip or the Asia Screening Array, and have undergone stringent quality controls and imputation. Using this cohort, we found 33 genetic loci associated with IBD beyond genome-wide significance, among which 17 had never been reported in IBD genetic studies in EAS. Two of these new IBD-associated loci in EAS are implicated by coding variants in *RUNX3* (P=2.9e-8) and *ADAP1* (P=2.8e-8). *RUNX3* is a transcription factor playing an important role during the development of T cells and regulating TGFÎ² signaling. Runx3 knockout mice spontaneously develop IBD. *ADAP1* is involved in the BCR Signaling Pathway. We also found a pleiotropic variant (P=1.1e-8), located in the intron of *GTF21*, associated with both IBD and systemic lupus erythematosus. Across the genome, we found common variants underlying IBD genetic risk have similar genetic effects between EAS and EUR ancestries, with a genetic correlation of 0.95. We also found the odds ratios of IBD putative causal variants highly consistent across ancestries (slope: 0.9). Despite the overall consistency and in line with previous reports, we found a few loci showing clearly different genetic effects in EAS vs EUR, including *TNFSF15* and *IL23R*, suggesting gene-environment interactions modifying the genetic risks across populations. Encouraged by the overall consistency of genetic effects across populations, we performed a fixed-effect meta-analysis with the latest European IBD GWAS and identified over 50 new loci associated with IBD. In summary, we have demonstrated the value of including diverse ancestries in IBD genetics research through building and leveraging the largest IBD genetics cohort of non-European ancestry. Through joint and comparative analyses with European IBD GWAS, we have identified over 50 new IBD genetic loci and revealed important insights into the genetic IBD epidemiology across populations.

PrgmNr 1454 - High-resolution genomic architecture of COVID-19 severe disease using multi-ethnic whole genome sequencing data

View session detail

Author Block: **A. Kousathanas**[1], E. Pairo-Castineira[2,3], A. Stuckey[1], C. Odhams[1], S. Walker[1], S. Clohisey[2], K. Rawlik[2], A. Fawkes[4], D. J. Rhodes[1], A. Siddiq[1], P. Goddard[1], S. Donovan[1], ISARIC4C Investigators, Genomics England Research Consortium, GenOMICC Investigators, C. P. Ponting[3], K. Rowan[5], L. Murphy[4], P. J.M. Openshaw[6], M. G. Semple[7], A. Rendon[1], R. H. Scott[1], A. Law[2], L. Moutsianas[1], M. Caulfield[1,8], K. J. Baillie[2,3,9]; [1]Genomics England, London, United Kingdom, [2]Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom, [3]MRC Human Genetics Unit, The MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, United Kingdom, [4]Edinburgh Clinical Research Facility, Western General Hospital, University of Edinburgh, Edinburgh, United Kingdom, [5]Intensive Care National Audit & Research Centre, London, United Kingdom, [6]National Heart and Lung Institute, Imperial College London, London, United Kingdom, [7]NIHR Health Protection Research Unit for Emerging and Zoonotic Infections, Institute of Infection, Veterinary and Ecological Sciences University of Liverpool, Liverpool, United Kingdom, [8]William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom, [9]Intensive Care Unit, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom
Disclosure Block: **A. Kousathanas:** None.
With the aim of exploring host genetic factors underlying COVID-19 severity, the Genetics Of Mortality In Critical Care consortium (GenOMICC) and Genomics England are recruiting participants with severe and mild COVID-19. Our previous work in the GenOMICC consortium revealed therapeutically-targetable variants underlying life threatening COVID-19 (Pairo-Castineira et al. 2020). Working with the global Host Genetics Initiative consortium (HGI), we have shown that the strongest genetic signals are present in the critically ill population. However, most previous work, including our own, has used genotyping arrays and largely focused on a subset of common genetic variation. In order to facilitate a high-resolution analysis of the genome and exploration of all types of genetic variation underlying disease risk, we are performing whole-genome sequencing (WGS) of up to 35,000 participants with severe and mild COVID-19. To increase statistical power, we are combining these data with WGS data generated through the 100,000 Genomes Project from a wide range of ancestral backgrounds. Recruitment is ongoing with successive rounds of analysis and data freezes being made available in a cloud-based research environment to facilitate international research efforts. Our latest data freeze consists of over 10,000 individuals with COVID-19, including the largest WGS cohort of critically ill COVID-19 patients assessed to-date. GWAS analysis of data on a freeze of 4,677 patients, replicated genetic associations previously identified by us, HGI and other groups on chromosomes 3,12,19, and 21. The WGS data enabled precise fine-mapping that narrowed down the 3p21.31 association cluster to two independent signals. Trans-ancestry GWAS analysis showed that one of the association signals on 3p21.31 can be independently replicated in multiple ethnicities but with significantly different effect sizes, underscoring the heterogenous impact of genetic ancestry on COVID-19. Rare variant burden analyses have yet to identify robustly any genes associated with severe COVID-19, which is consistent with other recent findings (Kosmicki et al. 2021). Our study provides the first comprehensive investigation of the contribution of rare and common genetic variation on COVID-19 disease outcomes using large scale WGS data, which will be instrumental in elucidating the genomic architecture of the disease and inform therapeutic interventions.

PrgmNr 1455 - A *de novo* paradigm for male infertility

View session detail

Author Block: **M. Xavier**[1], M. S. Oud[2], R. Smits[3], H. E. Smith[1], F. K. Mastrorosa[1], G. S. Holt[1], B. J. Houston[4], P. F. de Vries[2], B. K. S. Alobaidi[1], L. E. Batty[1], H. Ismail[1], J. Greenwood[5], H. Sheth[6], A. Mikulasova[1], G. Astuti[7], C. Gilissen[8], K. McEleny[9], H. Turner[10], J. Coxhead[11], S. Cockell[12], D. Braat[13], K. Fleischer[3,14], K. Dâ□□Hauwers[15], E. Schaafsma[16], GEMINI Consortium, L. Nagirnaja[17], D. F. Conrad[17], C. Friedrich[18], S. Kliesch[19], K. I. Aston[20], A. Riera-Escamilla[21], C. Krausz[22], C. G. Gonzaga-Jauregui[23], M. F. Santibanez-Koref[24], D. J. Elliott[1], L. E. Vissers[25], F. TÃ¼ttelmann[18], M. O'Bryan[26], L. Ramos[3], G. W. van der Heijden[2,3], J. A. Veltman[27]; [1]Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, [2]Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboudumc, Nijmegen, Netherlands, [3]Department of Obstetrics and Gynaecology, Radboudumc, Nijmegen, Netherlands, [4]School of BioSciences, Faculty of Science, The University of Melbourne, Melbourne, Australia, [5]Department of Genetic Medicine, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom, [6]FRIGE's Institute of Human Genetics, Ahmedabad, India, [7]Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboudumc, Nijmegen, Netherlands, [8]Radboud University Medical Center, Nijmegen, Nederland, Netherlands, [9]Newcastle Fertility Centre, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom, [10]Department of Cellular Pathology, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom, [11]Genomics Core Facility, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, [12]Bioinformatics Support Unit, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, [13]Department of Obstetrics and Gynaecology, Radboudumc, Newcastle upon Tyne, United Kingdom, [14]TFP Center of Reproductive Medicine, DÃ¼sseldorf, Germany, [15]Department of Urology, Radboudumc, Nijmegen, Netherlands, [16]Department of Pathology, Radboudumc, Nijmegen, Netherlands, [17]Division of Genetics, Oregon National Primate Research Center, Oregon Health & Science University, Beaverton, OR, [18]Institute of Reproductive Genetics, University of MÃ¼nster, MÃ¼nster, Germany, [19]Centre of Reproductive Medicine and Andrology, Department of Clinical and Surgical Andrology, University Hospital MÃ¼nster, MÃ¼nster,, Germany, [20]Univ of Utah, Salt Lake City, UT, [21]Andrology Department, FundaciÃ³ Puigvert, Universitat AutÃ²noma de Barcelona, Instituto de Investigaciones BiomÃ©dicas Sant Pau (IIB-Sant Pau), Barcelona, Spain, [22]Department of Biomedical, Experimental and Clinical Sciences â□□Mario Serioâ□□, University of Florence, Florence, Italy, [23]Regeneron Genetics Center, Tarrytown, NY, [24]Newcastle University, Newcastle upon Tyne, United Kingdom, [25]Radboudumc, Nijmegen, Netherlands, [26]School of BioSciences, Faculty of Science, The University of Melbourne, Parkville, Australia, [27]Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle-upon-Tyne, United Kingdom
Disclosure Block: **M. Xavier:** None.

De novo mutations (DNMs) are known to play a prominent role in many sporadic disorders with reduced fitness and genetic heterogeneity. Due to this strong effect on fitness, we hypothesize that DNMs play a prominent role in male infertility and explain a significant proportion of the genetic causes in this understudied disorder, where large-scale studies have to yet been published. In our study, we performed trio-based exome sequencing in a unique cohort of 185 males with unexplained cases of azoospermia or oligozoospermia and their unaffected parents. In total, 145 rare protein-altering *de novo* SNVs and 2 *de novo* CNVs were identified in these patients. Following a systematic analysis assessing mutational impact and protein function, 29 DNMs were classified as possibly causative of the male infertility phenotype observed in the affected patients. Additionally, a significant enrichment was detected in the number of Loss-of-Function (LoF) DNMs in LoF-intolerant genes ($p=1.00x10^{-5}$) and in predicted pathogenic missense DNMs in missense-intolerant genes

($p=5.01x10^{-4}$). Overall, a significant increase was found in the number of protein-protein interactions amongst genes affected by these DNMs ($p=2.35x10^{-2}$). Among the new candidate genes identified was *RBM5*, an essential regulator of male germ cell pre-mRNA splicing. Besides the patient carrying the DNM in *RBM5*, 6 additional infertile men were found carrying a distinct rare pathogenic missense mutation in *RBM5* in an international cohort of patients (n=2,506), a significant enrichment when compared to the number of mutations found in *RBM5* in a cohort of confirmed fertile men (n=5,784; p=0.03).Taken together, our results provide strong evidence for the role of DNMs in severe male infertility and identify a number of new candidate genes affecting human male fertility.

PrgmNr 1456 - Genomic and Transcriptomic-wide Analysis Identifies Novel Genetic Risk Loci and Prioritization of Therapies for Myasthenia Gravis

View session detail

Author Block: **R. Chia**[1], S. Saez-Atienzar[2], N. Murphy[1], A. ChiÃ²[3,4,5], C. Blauwendraat[6], the International Myasthenia Gravis Genomics Consortium, R. Roda[7], P. Tienari[8,9], H. Kaminski[10], R. Ricciardi[11], M. Guida[11], A. De Rosa[11], L. Petrucci[11], A. Evoli[12], C. Provenzano[13], D. Drachman[7], B. Traynor[14]; [1]National Institutes of Health, Bethesda, MD, [2]National Institute of Health, Bethesda, MD, [3]University of Turin, Turin, Italy, [4]Institute of Cognitive Sciences and Technologies, Rome, Italy, [5]Azienda Ospedaliero Universitaria Cittaì della Salute e della Scienza, Turin, Italy, [6]NIA, Bethesda, MD, [7]Johns Hopkins School of Medicine, Baltimore, MD, [8]Helsinki University Hospital, Helsinki, Finland, [9]University of Helsinki, Helsinki, Finland, [10]George Washington University, Washington DC, DC, [11]University of Pisa, Pisa, Italy, [12]Catholic University, Rome, Italy, [13]UniversitÃ Cattolica del Sacro Cuore, Rome, Italy, [14]National Inst Aging, Bethesda, MD

Disclosure Block: **R. Chia:** None.

Myasthenia gravis is a chronic autoimmune disorder caused by antibody-mediated destruction of receptors at the neuromuscular junction (NMJ) resulting in loss of communication between motor neuron and skeletal muscle. Genetics are likely to play a role as myasthenic patients have a higher risk of developing other autoimmune diseases. We performed a genome-wide association study (GWAS) and a transcriptome-wide association study (TWAS) to identify the genetic risks and candidate genes involved in disease etiology. The discovery cohort consists of 1,873 myasthenic patients with acetylcholine receptor antibodies and 36,370 healthy individuals, whereas 354 cases and 7,058 healthy individuals from the UK Biobank were included in the replication cohort. In addition to previous signals in *PTPN22*, *HLA-DQA1/HLA-B*, and *TNFRSF11A*, two novel signals in the acetylcholine receptor subunits genes, which are common antigenic target of the autoantibodies, were identified. From GWAS, the risk variant located in a promoter region of the cholinergic receptor nicotinic alpha 1 subunit (*CHRNA1*) gene on the forward strand increased risk by 1.57 ($p=3Â·07Ã10^{-8}$). TWAS identified two genes of which lower expression in skeletal muscle was predicted to increase disease risk: cholinergic receptor nicotinic beta 1 subunit (*CHRNB1,* $p=3Â·01x10^{-6}$, Z=-4Â·67*)* and epidermal growth factor receptor family of receptor tyrosine kinases (*ERBB2*, $p=5Â·63x10^{-7}$, Z=-5Â·00). Both *CHRNB1* and *ERBB2* are highly expressed at the NMJ and involved in the formation of functional acetylcholine complex. Thus, it is reasonable that lower expression of these proteins could impair neurotransmission. Onset-stratified analysis demonstrate that the genetic architecture is different between patients with early (CHRNA1, *CHRNB1* and *ERBB2* genes. In an unbiased approach, we identified several immunotherapies that may modify the disease progression based on their genetic profile. Our results demonstrate the power of genomics as a viable strategy for drug repurposing across diseases.

PrgmNr 1457 - Polygenic risk scores as a marker for lifetime epilepsy risk

View session detail

Author Block: **H. O. Heyne**[1,2], FinnGen, M. J. Daly[3,1], R. KÃ¤lviÃ¤inen[4]; [1]Institute for Molecular Medicine Finnland: FIMM, Helsinki, Finland, [2]Hasso Plattner Institute, Potsdam, Germany, [3]Massachusetts General Hospital, Boston, MA, [4]Kuopio University Hospital, Kuopio, Finland
Disclosure Block: **H.O. Heyne:** None.
Epilepsy affects approximately 1% of individuals worldwide. Making an epilepsy diagnosis is often difficult with estimates that up to 25% of epilepsy could initially be misdiagnosed. The SNP heritability of genetic generalized epilepsy is high (32%) and it has recently been shown that individuals with epilepsy also have elevated epilepsy polygenic risk scores (PRS). However, investigation how epilepsy PRS may predict epilepsy risk in an individual across lifetime or in individuals with unclear seizure events has so far been lacking.Here, we studied epilepsy PRS in detailed longitudinal electronic health records of > 269k Finns including ICD codes and drug purchases of over 50 years. Our dataset included 9660 individuals with epilepsy related diagnoses. We could confirm previous studies describing an elevated PRS for generalized epilepsy (PRS_{gen}) in individuals with generalized epilepsy. This was particularly high for juvenile myoclonic epilepsy, which could be because it represented the largest diagnosis group of the GWAS that was used to construct the PRS. Individuals with a top 10% PRS_{gen} in FinnGen had a more than doubled lifetime risk (p-value $1x10^{-11}$, hazard ratio 2.3) to develop generalized epilepsy compared to the bottom 90% PRS_{gen}. We further found that over half of individuals with specific diagnoses of generalized or focal epilepsy were initially diagnosed with unclear convulsions (R56.8) or unclear epilepsy (G40.9). Their PRS_{gen} was significantly higher than of those individuals who had only one unclear seizure event and did not later receive an epilepsy diagnosis. Specifically, individuals with a top 10% PRS_{gen} had a hazard ratio of 2.4 (p-value $1x10^{-4}$) to progress to generalized epilepsy after an unclear seizure event below the age of 40 compared to the bottom 90% PRS_{gen}. These results indicate a future potential for epilepsy PRS to help in predicting progression to epilepsy.

PrgmNr 1458 - Trans-ancestry GWAS meta-analysis of alcohol and tobacco addiction in 3.4 million individuals

View session detail

Author Block: **G. Saunders**, GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN), Trans-Omics for Precision Medicine (TOPMed); University of Minnesota, Minneapolis, MN
Disclosure Block: **G. Saunders:** None.
The use and abuse of nicotine and alcohol accounts for >100 million disability-adjusted life years across the globe, constituting one of the worldâ□□s leading public health problems. Despite this, the vast majority of genome-wide association studies (GWAS) thus far have been restricted to individuals of European ancestry, representing

PrgmNr 1459 - A Genealogical Estimate of Genetic Relationships to Improve Detection of Population Structure Over Time

View session detail

Author Block: **C. W. Chiang**, C. Fan, N. Mancuso; University of Southern California, Los Angeles, CA
Disclosure Block: **C.W. Chiang:** None.
The application of genetic relationships among individuals, characterized by a genetic relationship matrix (GRM), has far-reaching effects in genetic epidemiology. However, the current standard to calculate the GRM does not take advantage of linkage information and does not reflect the underlying genealogical history of the study sample. Here, we propose a coalescent-informed framework to infer the expected relatedness between pairs of individuals given an ancestral recombination graph (ARG) of the sample. This expected GRM (eGRM) is an unbiased and highly correlated estimate ($r^2 > 0.97$) of the latent pairwise genome-wide relatedness and maintains the mathematical properties of canonical GRMs. Through extensive simulations we show that the eGRM is robust when using genealogies inferred from incomplete genetic data, and can reveal the time-varying nature of population structure in a spatial sample. When applied to genotyping data from a population sample from Northern and Eastern Finland (N=2,644), we found that clustering analysis using the eGRM more accurately delineates population structure than would be possible using the standard GRM, and the temporal pattern of population structure in this sample is consistent with that of a recently diverged and expanded population. Taken together, our proposed estimator drastically shifts the notion of genetic relatedness from a variant-centric to a tree-centric world view, and will be widely applicable to genetic studies in understudied human or ecological samples where whole genome sequencing data or references might not be readily available.

PrgmNr 1460 - DNA methylation patterns underlying lifespan differences in mammals

View session detail

Author Block: **A. Haghani**, S. Horvath; University of California Los Angeles, Los Angeles, CA
Disclosure Block: **A. Haghani:** None.
The comparative cross-species analysis is a powerful tool to resolve the mysteries of evolution and phenotypic disparities among animals. This study describes the largest multi-species DNAm dataset that was collected by over 100 collaborators from Mammalian Methylation Consortium. This dataset includes over 10,000 DNA methylome data from multiple tissues of different age ranges of over 190 mammalian species. The network analysis of this dataset allowed us to identify co-methylation modules that relate to the individual (age, sex, tissue type) and species characteristics (e.g., phylogenetic order, maximum lifespan, adult weight). The unexpected correlation between DNA methylation and species was sufficiently strong to construct *phyloepigenetic* trees that parallel the phylogenetic tree. The analysis identified the epigenetic marks and the associated genes that relate to the maximum lifespan of mammals. Moreover, we could define novel epigenetic biomarkers of longevity that responded to gold standard anti-aging interventions in mice such as caloric restriction or growth hormone receptor knock outs. Our novel cross-species epigenetic analysis is a rich source of targets for future experimental studies of aging and longevity.

PrgmNr 1461 - Guaranteeing unbiasedness in selection tests based on polygenic scores

View session detail

Author Block: **J. Blanc**, J. J. Berg; University of Chicago, Chicago, IL
Disclosure Block: **J. Blanc:** None.
Population stratification is a well-studied problem in genome-wide association studies, leading to biases in the estimated strength of phenotypic association for individual genetic variants. In short, if environmental effects on the phenotype are correlated with ancestry gradients within a GWAS panel, any variant that is stratified along this ancestry gradient will receive a biased effect size estimate. While state of the art methods to correct for stratification are generally effective in reducing the number of significant false positive associations, even subtle biases in effect size estimates can accumulate across loci, leading to systematic biases in polygenic scores. In turn, these biases in the distribution of polygenic scores can lead to false positives in downstream analyses, such as tests for polygenic adaptation or other analyses of among group genetic differences. One approach is to attempt to be overly aggressive in controlling for stratification. However, there is currently no way to tell conclusively if confounding effects have been removed. A second approach is to conduct the GWAS in an evolutionarily diverged sample that is less likely to share population genetic structure with the test panel. This renders potential biases in the effect sizes irrelevant to the test, but comes at the cost of significantly reduced statistical power due to the issue of poor portability of polygenic scores across samples of divergent ancestry. Here using theory from population and statistical genetics, together with simulations, we show how this second approach can be generalized to panels that do share genetic structure, and that it is possible to guarantee the unbiasedness of selection tests without needing to guarantee that the effect sizes are fully unbiased. Specifically, if the researchers performing the GWAS also have access to the panel of test individuals and have identified the specific test to be performed ahead of time, then it is possible to compute a covariate to include in the GWAS, which will guarantee that the test is unbiased. Further, even when the test is not known prior to conducting the GWAS, our theoretical results provide a way to put a lower bound on the amount of stratification that would be needed to generate a positive signal of polygenic adaptation, and to assess whether the set of principal components that were included in the GWAS are sufficient to render the polygenic adaptation test unbiased. More generally, our results have implications beyond tests for selection as any analysis that attempts to quantify the correlation between polygenic scores and demographic or environmental variables is subject to the same type of stratification biases.

PrgmNr 1462 - A cross-disorder and evidence-based tiered ranking of candidate genes for neuropsychiatric disorders

View session detail

Author Block: **H. Shimelis**[1], M. T. Oetjens[2], M. Znidarsic[1], K. E. Wain[2], S. Myers[1], D. Ledbetter[1], C. L. Martin[2]; [1]Autism & Developmental Medicine Institute, Geisinger, Lewisburg, PA, [2]Geisinger, Lewisburg, PA

Disclosure Block: **H. Shimelis:** None.

Sequencing studies of cohorts with neuropsychiatric disorders (NPD) continue to identify many new NPD genes. These studies have revealed that categorical disorders, including intellectual disability (ID), autism spectrum disorder (ASD), and schizophrenia (SCZ), share genetic etiologies. Several databases have been created to catalogue rare variants in NPD probands, serving as a resource for researchers studying NPD. However, these databases generally use a single categorical NPD diagnosis in their approach, which is inconsistent with the known genetic overlap of these conditions and may underrepresent evidence for a given gene. To address this gap, the Developmental Brain Disorder (DBD) Gene Database (https://dbd.geisingeradmi.org/) uses a cross-disorder and tiered genotype-phenotype data mining approach to identify novel candidate genes and provide further evidence to genes previously implicated in NPD. Here, we present an update of the DBD Gene Database which contains data from 1172 studies published 2003-2020 representing 6481 individuals with pathogenic loss-of-function (pLOF) variants in 649 genes. pLOF and single gene copy number variants are curated from published sequencing studies across six NPD phenotypes: ID, ASD, epilepsy, attention deficit hyperactivity disorder, SCZ, and bipolar disorder. All genes are ranked into four tiers based on the number of cases with *de novo* pLOF variants: Tier 1, the strongest level of evidence, includes genes with three or more *de novo* pLOF variants; Tier 2, genes with two *de novo* pLOF variants; Tier 3, genes with one *de novo* pLOF variant; and Tier 4, genes with only inherited (or unknown inheritance) pLOF variants. Autosomal recessive (AR) genes are curated separately from LOF tier rankings. In our latest update from May 2021, 177 were ranked as Tier 1, 79 as Tier 2, 114 as Tier 3, 147 as Tier 4, and 132 as AR genes. To examine whether genes ranked as Tier 1 were novel or previously recognized as high-confidence genes in other databases, we compared Tier 1 genes to two NPD-related databases: SFARI gene and DDG2P (accessed May 2021). Of the 177 Tier 1 genes, 145 were listed as genes with the highest confidence of being NPD-related in at least one of the two databases while 32 were ranked lower (n=30) or not yet included (n=2). When we evaluated phenotypes of individuals with pLOF variants in Tier 1 genes, 96% (170/177) of genes were associated with 2 or more disorders. These results show that using a cross-disorder approach to NPD gene discovery increases the yield of genes with strong evidence of being NPD-related. Furthermore, our results show evidence for phenotypic heterogeneity in individuals with a pLOF variant in the same gene.

PrgmNr 1463 - Dysregulation of Canonical and Alternative Replication Protein A Complexes in Huntington Disease and Spinocerebellar ataxia Type 1 brains is associated with CAG instability and phenotype

View session detail

Author Block: **T. Gall-Duncan**[1], J. Luo[2], K. Fujita[3], C.-M. Jurkovic[4], V. Li[5], S. Tran[1], H. Tanaka[3], D. Leib[6], S. Lanni[2], N. Sato[7], G. B. Panigrahi[8], T. Prasolava[1], M. Curtis[9], L. Ranum[10], B. L. Davidson[11], E. Campos[1], A. R. La Spada[12], F. Boisvert[4], R. Faull[9], H. Okazawa[3], M. Wold[13], C. E. Pearson[14]; [1]Peter Gilgan Centre for Research and Learning, Toronto, ON, Canada, [2]The Hospital for Sick Children, Toronto, ON, Canada, [3]Tokyo Medical and Dental University, Tokyo, Japan, [4]UniversitÃ© de Sherbrooke, Sherbrooke, QC, Canada, [5]Toronto, ON, Canada, [6]Children's Hospital of Philadelphia, Philadelphia, PA, [7]Tokyo Med & Dental Univ, Tokyo, Japan, [8]Peter Gilgan Centre for Research and Learning, The Hospital for Sick Children, Toronto, ON, Canada, [9]University of Aukland, Aukland, New Zealand, [10]Univ Florida, Gainesville, FL, [11]The Children's Hospital of Philadelphia, Philadelphia, PA, [12]University of California, Irvine, Irvine, CA, [13]University of Iowa, Iowa, IA, [14]Hosp for Sick Children, Toronto, ON, Canada
Disclosure Block: **T. Gall-Duncan:** None.

Expansions of gene-specific CAG/CTG DNA repeats cause >15 neurodegenerative diseases, including Huntington Disease (HD) and spinocerebellar ataxia type 1 (SCA1). Inherited expansions continue to somatically expand as patients age, through a poorly understood mechanism. Larger expansions hasten disease onset and worsen severity and progression, so understanding the molecular processes of expansions is crucial to understanding pathogenesis. Somatic expansions may be regulated by tissue-specific expression of DNA repair proteins. Paradoxically, DNA repair proteins may exacerbate somatic expansions by incorrectly mediating repair of the expansion. We assessed the role of two single-strand DNA binding protein complexes in the expansion process. The canonical single-strand DNA binding complex in humans, replication protein A (RPA), composed of RPA1-RPA2-RPA3, is essential for DNA replication, repair, and recombination. Humans also express a poorly understood primate-specific alternative RPA complex (Alt-RPA) in which RPA4 replaces RPA2. Here we show RPA and Alt-RPA are differentially upregulated in HD and SCA1 patient brain regions, with Alt-RPA demonstrating up to 10-fold upregulation in the most affected brain tissues. In vitro repair of slipped-CAG structures, a DNA intermediate of expansions, shows that high concentrations of RPA enhance repair while high concentrations of Alt-RPA blocks repair. Coincidentally, while both Alt-RPA and RPA bind slipped-CAG structures, RPA efficiently melts slipped-DNAs while Alt-RPA does not. Conducting the first BioID interactomes for the RPA subunits we identified that RPA interacts with proteins known to protect against CAG-associated neurodegeneration, while Alt-RPA interacts with proteins which promote CAG-associated neurodegeneration and CAG expansions, including MSH3. We demonstrate that RPA overexpression completely inhibits somatic repeat expansions in vivo in the striatum of SCA1 mice, coinciding with reductions in biomarkers of CAG disease such as genome-wide DNA damage and mutant Ataxin-1 aggregation in striatal medium spiny neurons. Previously we demonstrated Rpa1 overexpression rescued motor phenotypes and ameliorated elevated DNA damage in cerebellar Purkinje neurons, in the same mice. Our new data demonstrate that somatic expansions in the striatum may contribute to SCA1 mouse phenotypes, and that RPA is an active player in suppressing CAG expansions and pathogenesis, effectively preventing somatic CAG expansions and molecular phenotypes through non-replication mechanisms. In contrast, Alt-RPA likely plays the opposite role by enhancing expansions and pathogenesis.

PrgmNr 1464 - Genotypic spectrum and its clinical implication in disorders with epilepsy in Indian population: A preliminary experience

View session detail

Author Block: **P. Majethia**[1], P. H. Somashekar[1], K. Nair[1], N. Jadhav[1], D. L. Narayanan[1], S. Mishra[1], A. Shrikiran[2], R. Bhat Y[2], S. Sharma[3], B. K. Praveen[4], B. L. Yatheesha[5], S. J. Patil[6], N. Kamath[7], S. Siddiqui[8], R. Kadavigere[9], S. L. Bielas[10], K. M. Girisha[1], A. Shukla[1]; [1]Medical Genetics, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, India, [2]Pediatrics, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, India, [3]Neurology Division, Department of Pediatrics, Lady Hardinge Medical College and Associated Kalawati Saran Children's Hospital, New Delhi, India, [4]Pediatrics, Father Muller Medical College Hospital, Mangalore, India, [5]Paediatric neurology, Dheemahi Child Neurology and Development center, Shimogga, India, [6]Pediatrics, Narayana Hrudayalaya Hospitals/ Mazumdar-Shaw Medical Center, Bangalore, India, [7]Pediatrics, Kasturba Medical College, Mangalore, Manipal Academy of Higher Education, Manipal, India, [8]Neuroimaging and Interventional Radiology, STAR Institute of Neurosciences, STAR Hospitals, Hyderabad, India, [9]Radio Diagnosis, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, India, [10]Department of Human Genetics and Department of Pediatrics, Division of Genetics, Metabolism, and Genomic Medicine, University of Michigan Medical School, Ann Arbor, MI
Disclosure Block: **P. Majethia:** None.
Introduction: Disorders with epilepsy are genetically and phenotypically heterogeneous. The rapid advancement of genetic testing has helped address these complexities and aided definitive diagnosis, precision medicine and genetic counseling in families with these disorders. **Methods:** We recruited 49 families (51 individuals) with neurodevelopmental disorder (NDDs) with epilepsy. Four individuals (four families) with clinically recognizable phenotypes underwent targeted genetic testing and 48 individuals (46 families) with clinically unrecognizable phenotypes and/or undiagnosed by targeted testing underwent genomic testing after detailed clinical evaluation. The implications on genetic counseling and therapy were evaluated in individuals with definitive diagnosis. **Results:** Our cohort comprises of 24 males (47%) and 27 females (53%) from 19 consanguineous (39%) and 30 non-consanguineous (61%) families. A definitive molecular diagnosis was achieved in 32 families with three being diagnosed by targeted testing and 29 by exome sequencing (ES). Twenty-nine monogenic disorders and one imprinting/microdeletion syndrome (Angelman syndrome) were identified. Of these, eight (25%) families had variants in seven genes causing developmental epileptic encephalopathies (*KCNQ2*, *STXBP1 UGDH*, *FGF13*, *AP3B2*, *FGF12*, *CYFIP2*). Seven (23%) families had variants in seven genes causing metabolic disorders (*HEXB, GCDH, GCSH, PNPT1, SHMT2, CARS2, SLC25A10*) of which six had variants in six nuclear encoded mitochondrial genes. Seven (23%) families had variants in 6 genes causing disorders with epilepsy as a core symptom (*SCN1A, KCTD7, NAXD, TRAPPC4, TPP1, KCNJ10*) and nine (29%) had variants in nine genes causing NDDs associated with epilepsy (*KMT2A, AP4S1, DYRK1A, RNASEH2C, DYNC1H1, GALNT2, ARID1B, ANKRD11, MECP2*). Biallelic variants were identified in 17 families (53%), *de novo* in 13 (41%), and heterozygous exonic deletion in one (3%) family. Seventeen (55%) of 31 disease-causing variants were novel. We also report variants in *SHMT2, SLC25A10*, and *FGF13* causing extremely rare disorders with less than 10 patients reported worldwide. Prenatal diagnosis was carried out in 19% (5/26) of families. Notably, ES had therapeutic implications in 50% of individuals (13/26) with a definitive diagnosis. **Conclusion:** The above-mentioned cohort is a part of an ongoing study of disorders with epilepsy. We herein describe the first cohort elucidating the genotypic spectrum and its clinical implications in disorders with epilepsies in Indian population.

PrgmNr 1465 - Identification of *NDEL1* as a novel gene for lissencephaly

View session detail

Author Block: **M.-H. TSAI**[1,2], H.-C. Ke[3], C.-Y. Fong[4], C.-C. Ng[4], K.-S. Lim[4], J.-W. Tsai[3]; [1]Kaohsiung Chang Gung Memorial Hospital, Kaohsiung, Taiwan, [2]Chang Gung University, Taoyuan, Taiwan, [3]National Yang-Ming Ciao-Tung Univeristy, Taipei, Taiwan, [4]University of Malaya, Kuala Lumpur, Malaysia

Disclosure Block: **M. Tsai:** None.

Background Lissencephaly (LIS) is a rare neurological disorder caused by abnormal neuronal migration. More than 20 LIS genes have been reported so far, most of them are functionally associated with the cytoskeleton. In a recent large study, more than 80% of patients with lissencephaly can have a pathogenic variant in reported genes. The genetics of lissencephaly in Asian patients is relatively under-explored. In addition, novel genes remain to be identified in unsolved cases. Method We recruited 33 lissencephaly patients and their family members. Candidate gene Sanger sequencing or whole exome sequencing (WES) was performed to search for known genetic causes. Trio WES study was used to study unsolved cases. In utero electrophoresis model was performed to investigate the effect of NDEL1 mutant on neuronal migration. Results In total, 33 patients with lissencephaly were enrolled in this study. 20/33 (60.6%) had a pathogenic variant in reported genes, including 7 DCX, 2 CEP85L, 2 DYNC1H1, 2 MAST1, 2 chromosome 17 deletion, 1 PAFAH1B1, 1 WDR62, 1 TUBA1A, 1 GRIN1, and 1 BICD2. WES Trio study was performed on the remaining 13/33 (39.3%) patients. One patient with a de novo NDEL1 missense variant was identified. NDEL1 is known to interact with LIS1 (PAFAH1B1) and required for microtubule organization and anchoring microtubule to the centrosome. Our preliminary functional study demonstrated that NDEL1 knock down and mutant both impaired neuronal migration. Conclusion The currently known lissencephaly genes are accounted for ~60% of our Asian cohort. We reported that NDEL1 is a novel gene for lissencephaly in human. NDEL1 dysfunction impaired neuronal migration during brain development.

PrgmNr 1466 - *EFEMP1* rare variants cause juvenile-onset open angle glaucoma in families from the Philippines

View session detail

Author Block: **J. L. Wiggs**[1,2], R. Collantes[1,3], M. S. Delfin[3], B. Fan[1], J. M. R. Torregosa[4], C. Siguan-Bell[4], N. Florcruz[5,6], J. D. Martinez[6], B. Masna-Hidalgo[7], V. T. Guzman[7], J. Anotado-Flores[7], F. D. Levina[8], S. C. Hernandez.[9], A. A. Collantes[10], M. Sibulo[11], S. Rong[1]; [1]Harvard Med Sch, MEEI, Boston, MA, [2]Broad Institute, Boston, MA, [3]Department of Ophthalmology, Manila Doctors Hospital, Manila, Philippines, [4]Department of Ophthalmology, Cebu Institute of Medicine, Cebu Velez General Hospital, Cebu City, Philippines, [5]Department of Ophthalmology and Visual Sciences, University of the Philippinesâ□□Philippine General Hospital, Manila, Philippines, [6]Glaucoma Service, Department of Ophthalmology, East Avenue Medical Center, Quezon City, Philippines, [7]Department of Ophthalmology, Western Visayas Medical Center, Iloilo City, Philippines, [8]Department of Ophthalmology, Jose R. Reyes Memorial Medical Center, Manila, Philippines, [9]UmeÃ¥ University, UmeÃ¥, Sweden, UmeÃ¥, Sweden, [10]College of Medicine, Our Lady of Fatima University, Valenzuela, Philippines, [11]Faculty of Medicine and Surgery, University of Santo Tomas, Manila, Philippines
Disclosure Block: **J.L. Wiggs:** Consultant/Consulting Fees/Other Remuneration; Aerpio, Allergan, Editas, RegenXbio, Avellino, Maze.

Glaucoma is the leading cause of irreversible blindness world-wide and exhibits both Mendelian (childhood onset) and complex (adult onset) inheritance. Ten genes are known to cause childhood glaucoma, but in total these only account for disease in approximately 20% of patients. Because most of the childhood glaucoma genes have been identified in European Caucasian populations, to facilitate novel gene discovery we have studied affected families from various geographic regions including the Philippines. As part of this effort, we used exome sequencing to evaluate 14 Filipino juvenile open-angle glaucoma (JOAG) families and identified 3 independent families (35, 2 and 27 members) with rare *EFEMP1* variants (p.N80Y, p.R477C and p.Ter494Glnext*29) co-segregating with disease. None of the rare variants are present in gnomAD and all modify highly conserved amino acids. Affected variant carriers (N= 34) exhibited severe disease with average age of disease onset of 16 years (range 3-43) and 76% developing blindness. Common SNPs near *EFEMP1* have been associated with adult-onset glaucoma (POAG) and a low frequency *EFEMP1* variant (p.R140W; MAF 0.0008%) may contribute to disease risk in one POAG family. Interestingly, a single *EFEMP1* missense allele (p.R345W) is known to cause Malattia Leventinese (ML), an inherited retinal degeneration. *EFEMP1* is an extracellular matrix protein with ocular expression similar to Myocilin, another extracellular matrix protein known to cause JOAG through a mechanism involving protein misfolding and endoplasmic reticulum aggregation. To determine if *EFEMP1* variants exhibit similar effects, we transfected COS7 cells with vectors expressing the three novel *EFEMP1* variants as well as variants associated with POAG and ML. We showed that all three variants found in JOAG patients caused significant intracellular protein aggregation compared to wild type and also the variants associated with the other phenotypes (p.R140W and p.R345W). These results suggest that rare coding *EFEMP1* variants can cause JOAG through a mechanism involving protein aggregation and that the extent of intracellular protein aggregation and retention appears to be the basis for the observed phenotype spectrum. This is the first report of *EFEMP1* variants causing early-onset glaucoma and we show that *EFEMP1* variation appears to be a relatively common cause of childhood glaucoma in these Filipino families. These results underscore the value of ethnically diverse populations for comprehensive detection of disease-causing mutations.

PrgmNr 1469 - *Genetics Adviser: The development, usability and acceptance testing of a patient-centered digital health application to support clinical genomic testing*

View session detail

Author Block: **M. Clausen**[1], R. Kodida[2], S. Shickh[3], E. Reble[4], C. Mighton[5], D. Hirjikaka[2], J. Sam[6], S. Krishnapillai[2], E. Adi-Wauran, G. Feldman[2], E. A. Glogowski[7], S. Shastri-Estrada[2], A. Scheer[2], E. Seto[8], C. T. Shuman[9], N. Baxter[10], A. Eisen[11], C. Elser[12], R. Kim[13], J. Lerner-Ellis[12], J. C. Carroll[14], K. A. Schrader[15], H. Faghfoury[16], Y. Bombard[14]; [1]209 Victoria Street, Toronto, ON, Canada, [2]St. Michael's Hospital, Toronto, ON, Canada, [3]St. Michael's Hospital and University of Toronto, Toronto, ON, Canada, [4]St. Michael's Hospital, Toronto, Canada, [5]St. Michael's Hospital & University of Toronto, Toronto, ON, Canada, [6]St. Michael's Hospital, North York, ON, Canada, [7]Mem Sloan Kettering Cancer Ctr, New York, NY, [8]University of Toronto, Toronto, ON, Canada, [9]Hosp for Sick Children, Toronto, ON, Canada, [10]University of Melbourne, Melbourne, Australia, [11]Sunnybrook Hospital, Toronto, ON, Canada, [12]Mount Sinai Hospital, Toronto, ON, Canada, [13]University Health Network/Mt Sinai Hospital, Toronto, ON, Canada, [14]Univ of Toronto, Toronto, ON, Canada, [15]Univ British Columbia, Vancouver, Canada, [16]Fred a Litwin and family Centre in Genetic Medicine, Toronto, ON, Canada

Disclosure Block: **M. Clausen:** None.

Background: Increasing demand for genomic testing coupled with existing workforce shortages in clinical genetics has placed unsustainable pressure on the standard models of care. Patient-facing digital health applications can empower patients and provide sustainable and scalable clinical solutions to address this gap. **Aim:** To transform our original Genomics ADvISER decision aid into a comprehensive patient-centered digital health application that will deliver education, counseling, and return of results for patients undergoing various form of genomic testing. **Methods:** Driven by user-centred design principles, we developed and conducted usability and acceptance testing of the application using an iterative, mixed-methods process consisting of: 1) consultations with an advisory board of providers and patients; 2) analysis of qualitative interviews with prior patients who used the original Genomics ADvISER decision aid; 3) creation of a digital wireframe prototype; 4) usability testing of the prototype; and 5) acceptability testing of the final digital application. ***Results:*** **Prototype development:** We created a new digital application, called the â□□Genetics Adviserâ□□, building on our original â□□Genomics ADvISERâ□□ that incorporated feedback from our advisory board, patients, genetics experts and the general public. The Genetics Adviser is designed to be easily adaptable to the needs of different types of patients, test modalities, and results. It consists of a pre-test module that focuses on: education, values, FAQs, patient stories/vignettes and the selection of results. It also includes a post-test check-in module to support users while they wait for results and a function that allows clinicians to upload results for patients to review before or after clinical consults. **Usability testing:** We conducted 25 usability tests with patients, the general public and genetics practitioners (15/25 female; mean age 41 years; 5/25 diagnosed with cancer). Participants were enthusiastic about the application, found it easy to navigate and comprehend. Participants recommended clarifying content and outlining the purpose of tasks. **Acceptance testing:** The application is currently undergoing qualitative and quantitative acceptability testing with a sample of patients and the general public (n=20). The final application will then be evaluated in a RCT with patients undergoing genomic testing. **Conclusions:** We created and tested an interactive, patient-centered application to optimize delivery, access and quality of care for pre- and post-test genomic testing, counselling, and return of results adaptable to any testing platform and setting.

PrgmNr 1470 - The cost of good health: Poverty association with differential gene expression

View session detail

Author Block: **N. Arnold**[1,2], J. Resztak[2], A. Alazizi[2], S. Dubaisi[2], R. J. Thorpe Jr[3], N. Noren Hooten[4], M. K. Evans[4], D. F. Dluzen[5,1], R. Pique-Regi[6], F. Luca[6]; [1]Morgan State University, Baltimore, MD, [2]Wayne State University, Detroit, MI, [3]John Hopkins University, Baltimore, MD, [4]National Institute on Aging, National Institutes of Health, Baltimore, MD, [5]National Institutes of Health, Silver Spring, MD, [6]Wayne State Univ, Detroit, MI

Disclosure Block: **N. Arnold:** None.

PURPOSE Psychosocial factors exert a powerful influence on health status and contribute to health disparities among marginalized populations. For example, overall life expectancy at birth throughout the United States tracks with poverty level, educational attainment, economic security and other upstream social determinants of health. Socioeconomic status (SES) and psychosocial factors are documented to affect gene expression in peripheral blood mononuclear cells, suggesting a molecular mechanism for some health disparities. Here we investigated the effects of poverty among Baltimore City residents participating in the Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS). **METHODS** We examined 239 participants of the HANDLS cohort study in Baltimore whose reported household income was either above or below the 125% federal poverty line for 2004. This population sample was composed of 119 African Americans and 120 white, for a total of 119 men and 120 women. We performed RNA sequencing in peripheral blood mononuclear cells to assess differential gene expression patterns associated with poverty. **RESULTS** We identified 15 genes differentially expressed when testing for poverty while controlling for race, sex, and age. When focusing on women, we found that individuals living in poverty had increased expression for 9 genes and decreased expression for 11 genes compared to individuals living above the poverty line. GSEA identified an enrichment for Herpes simplex virus infection pathway and B cell mediated immunity in genes differentially expressed in women living in poverty. **CONCLUSIONS** Our study suggests that poverty status influences gene expression in the immune system. Improving health outcomes for at-risk populations is achievable by understanding the link between poverty and identifiable biological mechanisms that influence disease.

PrgmNr 1471 - The impact of clinically relevant CNVs in the general population - the health consequences and personalized management of undiagnosed adult CNV carriers in the Estonian biobank

View session detail

Author Block: **M. NÃµukas**[1], M. Palover[1], M. Lepamets[1], L. Pizzo[2], K. Lehto[1], A. Reigo[1], H. Alavere[1], L. Leitsalu[1], I. Xenarios[3,4], M. Metspalu[5], A. Reymond[6], S. Girirajan[2], N. TÃµnisson[1,7], K. MÃ¤nnik[3]; [1]Estonian Biobank, Institute of Genomics, University of Tartu, Tartu, Estonia, [2]Biochemistry & Molecular Biology, and Anthropology, Pennsylvania State University, University Park, PA, [3]Health2030 Genome Center, Geneva, Switzerland, [4]Lausanne University Hospital, Lausanne, Switzerland, [5]University of Tartu, Tartu, Estonia, [6]Univ Lausanne, Lausanne, Switzerland, [7]Department of Clinical Genetics in Tallinn, Tartu University Hospital, Tartu, Estonia
Disclosure Block: **M. NÃµukas:** None.
The role of CNVs is well established in children with neurodevelopmental disorders (NDDs). However, more research is needed towards systematic understanding on how: i) CNVs affect health in the adult population and ii) to responsibly disclose findings to CNV carriers at high genetic risk for complex disorders.
We screened the cohort of Estonian Biobank (EstBB; n=132,770) for 81 recurrent CNV regions associated with susceptibility to NDDs. In the first stage, we used the â□□genotype-firstâ□□ approach to fetch health data from the EstBB, linked electronic health registries (EHRs) and mapped each ND-CNV with their co-occurring disease traits. In the second stage, we selected 10 ND-CNVs as a paradigm to return of genetic risk data and analysis of at-risk individualsâ□□ experience and the impact of disclosed genetic finding on their health support.
Our results show that the prevalence of CNVs associated with susceptibility to NDDs in the EstBB is 2.6% (n=3,404). Further prioritization of CNVs listed in the DECIPHER database suggested a population prevalence of 0.5% (n=710) for clinically well-established CNV syndromes. According to EHRs, nearly half of them have previously documented neurological or mental and behavioural problems. Notably, only 4 out of 710 are aware of their genetic diagnosis.
Our results along with reports by others confirm that CNVs associated with NDDs are cumulatively common, but still understudied health problem in general population. This work was supported by the Jacobs Foundation Research Fellowship (Dr MÃ¤nnik), Swiss National Science Foundation grant (Dr A.R) and the Estonian Research Council grant (Dr TÃµnisson).

PrgmNr 1472 - Uptake of cancer risk management strategies among women who undergo cascade genetic testing for breast cancer susceptibility: How do they compare to non-cascade testers?

View session detail

Author Block: **S. Makhnoon**[1], L. Brooke[2], K. Mattie[2], R. Volk[1], G. Grana[3], B. Arun[1], S. K. Peterson[1]; [1]The University of Texas MD Anderson Cancer Center, Houston, TX, [2]MD Anderson Cancer Center at Cooper University Health Care, Camden, NJ, [3]Cooper Hlth System, Voorhees, NJ
Disclosure Block: **S. Makhnoon:** None.
Purpose: A primary rationale for cascade testing for hereditary breast and ovarian cancers (HBOC) is to offer cancer risk management options, including bilateral mastectomy (BLM), salpingo-oophorectomy (BSO), and intensified screening, to persons found to have an inherited predisposition to the disease. At-risk relatives of probands with a pathogenic variant have familial experiences with cancer genetic testing which may shape health beliefs and subsequent health behaviors. Yet, little is known about their genetically-informed cancer risk-reduction behaviors, and long-term outcome data among cascade testers are lacking. This study evaluated: (1) the uptake and timing of cancer preventive surgical and screening strategies between cascade and non-cascade testers, and (2) the association between uptake of cancer risk management strategies and proband characteristics. Methods: Medical records were abstracted for all unaffected women with pathogenic variants in HBOC-associated genes from two cancer hospitals with at least one year of follow-up to compare uptake of surgery and screening between cascade and non-cascade testers. Results: 341 women underwent post-test genetic counseling between 2013 and 2019, of whom 253 were included in the analytic sample. Cascade testers (79.8%) were younger than non-cascade testers (mean=37.6 vs. 43.5 years, p=0.002) and most commonly had a parent with a pathogenic variant (39.1%) followed by siblings (21.3%), and multiple other relatives (20.8%). Women were predominantly non-Hispanic White (81.0%) and underwent testing for *BRCA1* (42.0%) or *BRCA2* (47.2%) variants. Among women ageâ¥40 years, 43% underwent BLM and 71.6% underwent BSO with no significant difference in uptake between cascade and non-cascade testers. Mean time to BSO among cascade testers was shorter among women ageâ¥40 vs Conclusion: Management uptake among cascade testers is high with rates comparable to unaffected *BRCA* positive women. A large proportion of women act on cascade test results and this represents a novel report of utilization of cancer management strategies

PrgmNr 1473 - Haplotyping SNPs for allele-specific gene editing of the mutant huntingtin allele using long-read sequencing

View session detail

Author Block: **L. Fang**[1], A. M. Monteys[1], A. DÃ¼rr[2], M. Keiser[1], C. Cheng[1], A. Harapanahalli[1], P. Gonzalez-Alegre[1,3], B. L. Davidson[4], K. Wang[5]; [1]Children's Hospital of Philadelphia, PHILADELPHIA, PA, [2]Sorbonne UniversitÃ©, Paris, France, [3]University of Pennsylvania, Philadelphia, PA, [4]The Children's Hospital of Philadelphia, Philadelphia, PA, [5]Childrenâ□□s Hospital of Philadelphia, Philadelphia, PA
Disclosure Block: **L. Fang:** None.

Huntington's disease (HD) is an autosomal dominant neurodegenerative disease caused by CAG repeat expansions in the huntingtin (HTT) gene. Although trials of disease-modifying treatments are now on the horizon, the clinical care is focused on symptom management. We previously reported allele-specific deletion of the mutant HTT by CRISPR/Cas9 in a mouse model and human cell lines. Allele selectivity is achieved by targeting heterozygous SNPs that create or eliminate a Protospacer Adjacent Motif (PAM). However, given the lack of knowledge on haplotype structure in HD populations, a comprehensive analysis of all potential targeting sites is lacking and the optimal personalized editing strategy for HD individuals is unknown. To address this, we developed a multiplexed targeted long-read sequencing approach to sequence a 10.4 kb genomic region flanking exon-1 of HTT and created necessary computational tools (AmpBinner and AmpRepeat) to de-multiplex the data, detect repeats, and phase the reads. We applied this approach to two independent HD cohorts (974 individuals from the US and France), detected SNPs, analyzed haplotypes and potential editing sites for various enzymes. In the haplotype structure analysis, we showed potential founder effects in unrelated HD individuals from different continents. Based on the haplotype analysis, 23% of HD individuals of European ancestry can be edited by targeting one SNP (rs2857935). Up to 56% HD individuals of European ancestry can be potentially edited by combinatorial targeting of multiple SNPs. Our results provide the first haplotype map of the region surrounding exon-1 of HTT in HD cohorts. Our workflow can be applied to other repeat expansion diseases to facilitate allele-specific gene editing.

PrgmNr 1474 - A mouse model recapitulating Bruck Syndrome provides insight into the role of *Plod2* in bone and cartilage

View session detail

Author Block: **A. Kot**[1], I. Duran[1,2], J. H. Martin[1], D. Wachtell[1], M. Weis[3], D. R. Eyre[3], J. Zieba[1], D. Krakow[1]; [1]University of California, Los Angeles, Los Angeles, CA, [2]University of Malaga, Malaga, Spain, [3]University of Washington, Seattle, WA

Disclosure Block: **A. Kot:** None.

Bruck Syndrome is an autosomal recessive form of osteogenesis imperfecta characterized by bone fragility, short stature, and joint contractures resultant from biallelic mutations in either the genes encoding *PLOD2* or *FKBP10*. *PLOD2* encodes LH2 which is a component of an endoplasmic reticulum (ER) complex involved in type-I collagen telopeptide hydroxylation. Bruck Syndrome has features distinct from other forms of osteogenesis imperfecta, such as joint contractures, with pathophysiology that is not well understood. Studies on a previously published *Plod2* knockout mouse were limited due to early embryonic lethality. To study the pathophysiology of musculoskeletal development in Bruck Syndrome, we generated a mouse model with a homozygous mutation identified in two independent families with recurrent fractures, scoliosis, and congenital contractures. The mutation, *Plod2* c.1559dupC predicting the amino acid change p.Val523Cysfs*7, was introduced using targeted homologous recombination. *Plod2* is alternatively spliced and translated into short (LH2a) or long (LH2b) isoforms. The c.1559dupC mutation is in exon 13A, specific to LH2b, and is predicted to lead to loss of the long form of the protein. *Plod2*$^{V523fs/V523fs}$ mice did not survive past P0. At E18.5, *Plod2*$^{V523fs/V523fs}$ mice were smaller, had bilateral forelimb and hindlimb contractures, and absent cervical spine curvature. LH2b mRNA was near absent in *Plod2*$^{V523fs/V523fs}$ mouse calvaria. LH2 protein levels in *Plod2*$^{V523fs/V523fs}$ osteoblasts were also decreased compared to wild type. Type I collagen C-telopeptide lysine residue hydroxylation was significantly reduced in *Plod2*$^{V523fs/V523fs}$ bone suggesting that crosslinking is negatively impacted, contributing to impaired bone properties. Compared to wild type, *Plod2*$^{V523fs/V523fs}$ growth plates showed increased length of the hypertrophic zone, increased hypertrophic chondrocyte volume, and diminished collagen production as detected by picrosirius staining suggesting a role for *Plod2* in chondrocyte development. This mouse model will aid in determining the function of *Plod2* in the musculoskeletal system and help uncover the molecular mechanisms underlying Bruck Syndrome.

PrgmNr 1476 - Mutations in DNA ligase III cause mitochondrial neurogastrointestinal encephalomyopathy

View session detail

Author Block: **M. Taniguchi**[1], E. Borona[2], C. Sanjiban[3], M. Tsutsumi[4], B. Francesca[2], G. Kellaris[5], C. Bergamini[6], H. Kurahashi[7], K. Kosaki[8], T. Toda[9], N. Katsanis[10], F. Duijkers[11], R. De Giorgio[12]; [1]Dep. Clin. Genet. Fujita Health University Hospital, Toyoake, Japan, [2]Dept, Medical and Surgical Sciences, Univ. of Bologna, Bologna, Italy, [3]Dept. Mol. Genet., Erasmus MC, Rotterdam, Netherlands, [4]Fujita Health University, Toyoake, Japan, [5]Center for Human Disease Modeling, Duke University, Durham, NC, [6]Dept. Pharmacy and Biotechnology, University of Bologna, Bologna, Italy, [7]ICMS, Fujita Health University, Toyoake, Japan, [8]Keio University School of Medicine, Tokyo, Japan, [9]The University of Tokyo, Tokyo, Japan, [10]Duke University, Durham, NC, [11]AMC, Utrecht, Netherlands, [12]University of Ferrara, Cona, Italy
Disclosure Block: **M. Taniguchi:** None.
Abnormal gut motility is a feature of several mitochondrial encephalomyopathies, and mutations in genes such as *TYMP* and *POLG*, have been linked to these rare diseases. The human genome encodes three DNA ligases, of which only one, ligase III (*LIG3*), has a mitochondrial splice variant and is crucial for mitochondrial health. We investigated the effect of reduced LIG3 activity and resulting mitochondrial dysfunction in seven patients from three independent families, who showed the common occurrence of gut dysmotility and neurological manifestations reminiscent of mitochondrial neurogastrointestinal encephalomyopathy. DNA from these patients was subjected to whole exome sequencing. In all patients, compound heterozygous variants in a new disease gene, *LIG3*, were identified. All variants were predicted to have a damaging effect on the protein. The *LIG3* gene encodes the only mitochondrial DNA (mtDNA) ligase and therefore plays a pivotal role in mtDNA repair and replication. *In vitro* assays in patient-derived cells showed a decrease in LIG3 protein levels and ligase activity. We demonstrated that the *LIG3* gene defects affect mtDNA maintenance, leading to mtDNA depletion without the accumulation of multiple deletions as observed in other mitochondrial disorders. This mitochondrial dysfunction is likely to cause the phenotypes observed in these patients. The most prominent and consistent clinical signs were severe gut dysmotility and neurological abnormalities, including leukoencephalopathy, epilepsy, migraine, stroke-like episodes, and neurogenic bladder. A decrease in the number of myenteric neurons, and increased fibrosis and elastin levels were the most prominent changes in the gut. Cytochrome *c* oxidase (COX) deficient fibres in skeletal muscle were also observed. Disruption of *lig3* in zebrafish reproduced the brain alterations and impaired gut transit *in vivo*. In conclusion, we identified variants in the *LIG3* gene that result in a mitochondrial disease characterized by predominant gut dysmotility, encephalopathy, and neuromuscular abnormalities.

PrgmNr 1477 - Significant burden of *de novo* damaging variants in novel genes in patients with congenital kidney malformations

View session detail

Author Block: **H. Milo Rasouly**[1], S. Krishna Murthy[2], S. Bedha[1], D. Fasel[3], M. Marasa[1], E. Fiaccadori[4], A. Materna-Kiryluk[5], G. Masnata[6], V. Tasic[7], M. Saraga[8], K. Kiryluk[3], G. Ghiggeri[9], S. Sanna-Cherchi[10], A. G. Gharavi[10]; [1]Columbia University, New York, NY, [2]Irving Medical Center , Columbia University, New York,, NY, [3]New York, NY, [4]Universita degli Studi di Parma Dipartimento di Medicina e Chirurgia, Emilia-Romagna, Parma, Italy, [5]Poznan University of Medical Sciences, Poznan, Poland, [6]Azienda Ospedaliera Brotzu, Cagliari,, Italy, [7]University Children's Hospital, Skopje, Macedonia, The Former Yugoslav Republic of, [8]University of Split, Split, Croatia, [9]Istituto Giannina Gaslini, Genova,, Italy, [10]Columbia Univ, New York, NY

Disclosure Block: **H. Milo Rasouly:** None.

Background: Congenital kidney malformations (CKM) are one of the most common cause of pediatric kidney failure. While multiple causative genes have been identified, they explain only 10-15% of cases. *De novo* variantsâ (DNV) analysis led to the identification of novel genes for congenital heart defects and neurodevelopmental disorders. We hypothesized that similar analyses could identify new CKM-causing genes. Methods: Patients enrolled through our study on genetics of kidney diseases underwent Exome sequencing (ES) or Genome sequencing (GS). ES was performed on 151 trios (i.e. affected child with CKM and unaffected parent), and GS on 53 additional trios.100 CKM trios were identified from the Deciphering Developmental Delay consortia (ES). The sequences were processed using a BWA/GATK 4.1 pipeline on the CAVATICA platform and annotated through Ensembl VEP. Custom bioinformatics pipelines were then used for data cleaning. Potential enrichment for DNV was analyzed with the denovolyzer package in R. Results: Out of the 304 trios analyzed, 10 cases (3%) had a DNV in one of the 172 genes known to be associated with dominant forms of kidney disease (3.2 fold-enrichment, p-value=1.41 x 10^{-3}). Those 10 cases included one protein-truncating variant (PTV) in *PAX2*, one in *HNF1B* and two PTV DNVs in *KAT6B*. Genome-wide, we observed a significant 2.8 fold-enrichment for PTV DNV (76 DNVs compared to 26.6 expected, p-value=4.3x10^{-15}) and a significant 1.3 fold-enrichment for missenses DNV (248 DNVs compared to 191.5 expected, p-value=5.1x10^{-5}). When constraining the analysis to genes highly expressed in nephron-progenitor cells at 18 weeks of gestation (Human fetus) and not known to be associated with kidney disease in Human, we observed significantly increased enrichment for PTV DNV (8 DNVs compared to 1.2 expected, OR=6.9, p=3.04 x 10^{-5}, list of 593 genes with a pLI>0.9 and a LOEUF>0.35), and for deleterious missense DNV (8 DNVs compared to 3.2 expected , OR= 2.51, p=0.016, list of 229 genes with a mis-z-score > 3.09). Gene-Set Enrichment Analysis uncovered a highly significant enrichment for genes down-regulated in ME-A cells (breast cancer) undergoing apoptosis in response to doxorubicin (FDR=1.5x10^{-11}) and in fibroblasts expressing mutant forms of ERCC3 after UV irradiation (FDR=5.5x10^{-8}).

Conclusions: We detected an excess of *de novo* variants in CKM, supporting potential pathogenetic mechanisms of disease. Globally, the DNV signal was partially driven by novel constrained genes that are highly expressed during early kidney development. Further gene-set analysis and replication in larger datasets may help pinpoint which genes are most likely driving this enrichment.

PrgmNr 1478 - Genome-wide genetic control of fetal placental genomics shows multiple associations with health and disease across the life course, informing the Developmental Origins of Health and Disease

View session detail

Author Block: **A. Bhattacharya**[1], A. N. Freedman[2], V. Avula[2], R. Harris[2], W. Liu[2], C. Pan[1], A. Lusis[3], R. M. Joseph[4], L. Smeester[2], H. J. Hartwell[2], K. C. K. Kuban[5], C. Marsit[6], Y. Li[2], T. O'Shea[2], R. C. Fry[2], H. P. Santos, Jr.[2]; [1]University of California at Los Angeles, Los Angeles, CA, [2]University of North Carolina at Chapel Hill, Chapel Hill, NC, [3]University of California, Los Angeles, Los Angeles, CA, [4]Boston University, Boston, MA, [5]Boston University School of Medicine, Boston, MA, [6]Emory University, Atlanta, GA
Disclosure Block: **A. Bhattacharya:** None.
The placenta is the master regulator of the intrauterine environment and is central to the Developmental Origins of Health and Disease (DOHaD). Studies show that fetal genetics and placental genomics can influence child health traits. An integrative analysis of genetics, placental genomics, and child health traits has not been done and would yield insight into the DOHaD hypothesis. Recently, we developed Multi-Omic Strategies for Transcriptome-Wide Association Studies (MOSTWAS), which uses mediation analysis to scan variants genome-wide, detect gene-trait associations (GTAs), and develop hypotheses for trait-associated gene regulation. Here, using genetic, transcriptomic, and methylomic data from the Extremely Low Gestational Age Newborn (ELGAN) Cohort Study (N = 272), we applied MOSTWAS to train genetic models of expression of all genes on the ELGAN RNA-seq panel, 2,994 of which showed strong in- and out-sample accuracy. With these models, we conducted TWAS for 40 traits from 5 categories (autoimmune, metabolic, cardiovascular, perinatal, neuropsychiatric) and identified 264 GTAs across 176 TWAS genes and potential transcription (TFs) or epigenomic factors regulating their expression. Of the 176 genes, 50 were associated with multiple traits, many not genetically correlated (e.g., *IDI1* with BMI and schizophrenia). Genetically-regulated placental expression, in aggregate, explained significant portions of three neonatal traits (total puberty growth, childhood BMI, start of puberty) at 5-8% of total SNP heritability. In addition, 91 GTAs showed significant associations through distal variants, with many mediated through 8 TFs associated with multiple TWAS genes. For example, *EPS15*, a maternally imprinted placental TF associated with fetal growth, showed a negative association with waist-hip ratio (WHR) and was negatively associated with two genes: *SPATA13* and *FAM214A*, both showing positive TWAS associations with WHR. In human placenta-derived trophoblasts, FANA silencing of *EPS15* led to upregulation of both *SPATA13* and *FAM214A*, providing evidence for placental TF regulation of these genes. Further transcriptomic and functional consequences of *EPS15* knockdown are under evaluation. Our study reveals potentially shared placental pathways associated with many traits across the life course. GTAs with traits across categories and different life periods suggest that placental dysregulation affects fundamental early-in-life traits, and these effects compound and manifest in later-in-life traits. Our work motivates increased sample sizes for early childhood trait GWAS and the placenta as a key tissue of study.

PrgmNr 1479 - Residual risk for clinically significant copy number variants in pregnancies with normal NIPS

View session detail

Author Block: **L. Sagi-Dain**[1], I. Maya[2], D. Brabbing-Goldstein[2], L. Basel-Salmon[2], L. Salzer[2]; [1]Carmel Medical Center, Haifa, Israel, [2]Rabin Medical Center, Petach Tikva, Israel
Disclosure Block: **L. Sagi-Dain:** None.
Background: Chromosomal microarray analysis (CMA) is the recommended first-tier in pregnancies with sonographic anomalies, while in low-risk pregnancies this test detects clinically significant copy number variants (CNVs) in about 1%. As the constantly growing wide use of non-invasive prenatal screening (NIPS) facilitates the detection of chromosomal aberrations, defining the residual risk for abnormal CMA following normal NIPS is of importance for informed decisions regarding prenatal testing and screening options. The objective of our study was to shed light on this issue. **Methods:** CMA results of all pregnancies undergoing amniocentesis between the years 2013-2021 in large hospital-based laboratory were collected. Pregnancies with major sonographic anomalies or multiple fetuses were excluded. Clinically significant (pathogenic and likely pathogenic) CNVs were divided into: 3-NIPS-detectable (trisomies 13, 18 and 21), 5-NIPS-detectable (including sex chromosome aberrations), 5-NIPS and common microdeletion-detectable (including 1p36.3-1p36.2, 4p16.3-4p16.2, 5p15.3-5p15.1, 15q11.2-15q13.1, and 22q11.2 deletions), and genome-wide NIPS-detectable (including variants >7Mb). The theoretical residual risk for clinically significant CNVs was calculated following exclusion of NIPS-detectable findings. **Results:** Of the 8,099 pregnancies, clinically significant CNVs were demonstrated in 70 pregnancies (1.4%). The residual risk following theoretically normal NIPS was 1.2% (1/85) for 3-NIPS, 0.92% (1/109) for 5-NIPS, 0.88% (1/113) for 5-NIPS including common microdeletions, and 0.82% (1/122) for genome wide NIPS. In the subgroup of 4,048 pregnancies with advanced maternal age, the residual risk for clinically significant CNVs following theoretically normal NIPS ranged from 1.3% (1/75) for 3-NIPS to 0.8% (1/122) for genome wide NIPS. In 3,187 pregnancies of women younger than 35 years, this residual risk ranged from 0.7% (1/145) for 3-NIPS to 0.5% (1/198) for genome wide NIPS. The residual risk was highest for the 559 pregnancies with abnormal serum screening and 305 pregnancies with soft markers - about 2.2% (1/45) for 3-NIPS and 2.0% (1/50) for genome wide NIPS. **Conclusions:** The residual risk of clinically significant CNVs in pregnancies without structural sonographic anomalies is appreciable, and depends on NIPS extent, maternal age, the results of biochemical screening and presence of soft markers. This knowledge is important for the patients, the obstetricians and the genetic counselors, in order to facilitate informed decisions regarding prenatal testing and screening options.

PrgmNr 1480 - A 100,000 Genome Project haplotype reference panel of 156,390 haplotypes and the improved imputation of UK Biobank

View session detail

Author Block: **S. Shi**[1], S. Rubinacci[2], S. Hu[3], L. Moutsianas[4], A. Stuckey[4], C. Cabrera[5,6], V. Cipriani[5,6], D. P. Smedley[5,6], M. J. Caulfield[7], S. R. Myers[1], J. L. Marchini[8]; [1]University of Oxford, Oxford, United Kingdom, [2]University of Lausanne, Ecublens, Switzerland, [3]Novo Nordisk Research Centre, Oxford, United Kingdom, [4]Genomics England, London, United Kingdom, [5]William Harvey Research Institute, London, United Kingdom, [6]Queen Mary University of London, London, United Kingdom, [7]Genomics England, Queen Mary, London, United Kingdom, [8]Regeneron Genetics Center, Tarrytown, NY
Disclosure Block: **S. Shi:** None.
The Genomics England (GEL) 100,000 genome project has sequenced over 85,000 genomes across England. By using high coverage whole-genome sequencing (WGS), this constitutes the largest human genetic variation resource ever collected in the UK, and represents a near-complete characterization of genetic variation in the population. We generated a GEL haplotype reference panel, comprising 341 million autosomal variants and 156,390 haplotypes from diverse ancestries. We exploit both the sample size and relatedness structure among individuals, 61.3% of whom possess at least one sequenced first-degree relative, to allow high-precision haplotypic phasing.
We used 1000 Genomes WGS data to assess the imputation performance across ancestries, and observe improvements in some populations. In samples of British origin the mean imputation r2 at 0.01% allele frequency is 0.45, 0.67 and 0.74 when using the HRC, TOPMed and GEL reference panel. In samples of South Asian origin the mean imputation r2 at 0.01% allele frequency is 0.04, 0.24 and 0.61 when using the HRC, TOPMed and GEL reference panel.
We used the GEL reference panel to impute the UK Biobank dataset, that was previously imputed at 39 million autosomal variants, using an HRC+UK10K reference panel. It results in a ~6 fold increase in the number of imputed variants. Mean information scores at imputed SNPs, from the GEL and HRC-UK10K reference panels, were 0.65 and 0.61 respectively. At low allele frequencies the differences were larger. For example, for SNPs with allele frequency between 0.01% to 0.1% mean information scores were 0.88 and 0.66 for the GEL and HRC-UK10K reference panels respectively. This translates into an appreciable boost in power to detect associations. The GEL-imputed UK Biobank dataset is being made available to all approved researchers of the UK Biobank.
We will also report results of experiments of examine the implications for fine mapping and burden association tests in the context of imputed GWAS for blood pressure and other traits.

PrgmNr 1481 - A powerful test of ancestral heterogeneity in the effects of gene expression on complex traits

View session detail

Author Block: **K. Knutson**[1], W. Pan[2]; [1]Division of Biostatistics, Univ Minnesota, Minneapolis, MN, [2]Univ Minnesota, Minneapolis, MN
Disclosure Block: **K. Knutson:** None.
The Transcriptome Wide Association Study (TWAS) is a widely used approach which integrates expression and GWAS data to study the role of cis-regulated gene expression (GEx) in complex traits. TWAS models GEx as a function of cis-eQTL genotypes. However, strong evidence suggests that the genetic architecture of GEx varies across populations. Furthermore, recent findings point to possible ancestral heterogeneity in the effects of GEx on complex traits, heterogeneity which may be amplified in TWAS by modeling GEx as a function of cis-eQTLs. We present a novel extension to TWAS which models heterogeneity in the effects of cis-regulated GEx which are correlated with ancestry. By jointly analyzing samples from multiple populations, our multi-ancestry TWAS framework can improve power to detect genes with shared expression-trait associations across populations through increased sample sizes, as compared to existing stratified TWAS approaches. Under our proposed model, we derive score tests for homogeneous, heterogeneous, and total GEx effects on a complex trait. Our preliminary simulations reveal conserved Type-I error rates and high power across a number of scenarios, holding promise for further simulations on larger simulated datasets. We apply our test to case-control genotypes from the Alzheimer's Disease Sequencing Project (ADSP) and publicly available prediction models from the Multi-Ethnic Study of Atherosclerosis (MESA) study. We identify a number of genes with suggestive heterogeneous effects between African American and Caucasian subjects on Alzheimerâ□□s Disease (AD), including *ASPHD2*, *SDSL*, *ZNF589*, *PGM2L1*, and *PPIL3*. Our preliminary application further identifies many putative AD risk genes which were not discovered through ancestry-stratified TWAS analyses, many of which have well established or biological plausible links to AD, including (but not limited to) *RTP4*, *ORAI2*, and *TOMM40L*. In forthcoming work, we will apply our test to a larger sample from ADSP, anticipating a greater number of significant findings due to the increased sample size. Additionally, we will apply our proposed test of heterogeneity to continuous endophenotypes from the UK Biobank (n â□□ 500,000), specifically considering a set of imaging derived phenotypes with strong associations to AD.

PrgmNr 1482 - Developing Trans-ethnic Polygenic Risk Scores Using Empirical Bayes and Super Learning Algorithm

View session detail

Author Block: **H. Zhang**[1], J. Zhan[2], J. Jin[3], J. Zhang[3], T. U. Ahearn[4], Z. Yu[5], J. O. Connell[2], Y. Jiang[2], B. Koelsch[2], 23andMe Research Team, X. Lin[1], M. Garcia-Closas[4], N. Chatterjee[3]; [1]Harvard T.H. Chan School of Public Health, Boston, MA, [2]23andMe Inc., Sunnyvale, CA, [3]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, [4]National Cancer Inst, Rockville, MD, [5]Broad Institute of MIT and Harvard, Cambridge, MA
Disclosure Block: **H. Zhang:** None.
Polygenic risk scores (PRS) are useful for predicting various phenotypes/outcomes; however, as most PRS have been developed with data generated in European Ancestry (EA) populations, performances of PRS are often poorer in non-EA populations, reflecting their degree of divergence from EA population.
To improve PRS performance in non-EA populations, we propose a novel method, Two-Dimensional Clumping and thresholding with Super Learning and Empirical Bayes (TDLD-SLEB), which takes advantage of both existing large GWAS from EA populations and smaller GWAS from non-EA populations. TDLD-SLEB uses a two-dimensional thresholding method to incorporate SNPs that have either effects in both the larger (e.g., EA populations) and the smaller (e.g., non-EA populations) target population or specific effects in the smaller population. It estimates effect sizes for SNPs in the target population using an Empirical Bayes method that borrows GWAS information across populations. Finally, it incorporates a super learning algorithm to combine series of PRS generated by various SNP selection thresholds for the target population.
Our simulation analyses mimicked real LD patterns using haplotype data of 1000 Genome Phase 3 for five ancestries. We considered various genetic architectures including different levels of negative selection and genetic correlation across ancestries. We found PRSs generated by TDLD-SLEB had significantly improved predictionary accuracy for non-EA populations in independent validation datasets, compared to single ethnic PRS, EUR derived PRS, or a weighted PRS that combines EUR and single ethnic derived PRS with weights selected to optimize prediction in the target population.
Using 23andMe data, we developed and validated population specific PRS for seven complex traits using GWAS data from Europeans (average N=2,442K), African American (average N=113K), Latino (average N=411K), East Asians (average N=94K), South Asians (average N=25K). We found TDLD-SLEB often led to large improvements in the performance of PRS compared to alternative methods for predicting traits in the African American population (average R^2 increased +277% compared to the weighted PRS method). For other ethnic groups, TDLD-SLEB also led to sometimes notable improvements in the performance of PRS, such as for the cardiovascular disease in the Latino population (AUC = 0.61 for TDLD-SLEB vs. AUC= 0.58 for the weighed PRS method).
In conclusion, we developed a computationally scalable and statistically efficient method for generating predictive PRS in non-European populations using GWAS datasets across diverse populations.

PrgmNr 1483 - Explainable and extendable machine learning models for identifying prognostic radiogenomic biomarkers from breast cancer multimodal imaging and genomic data

View session detail

Author Block: **Q. Liu**, P. Hu; University of Manitoba, Winnipeg, MB, Canada
Disclosure Block: **Q. Liu:** None.
Background: Radiogenomics is a field where medical images and genomic profiles are jointly analyzed to answer critical clinical questions. We proposed a novel framework to identify prognostic radiogenomic biomarkers from multi-modal breast cancer (BC) magnetic resonance imaging (MRI) and multi-omics data, which may serve as a substitute for genetic testing.
Methods: Bayesian tensor factorization (BTF) was used to extract the integrated multi-omics features from gene expression, DNA methylation, and copy number variation data of 762 BC patients. The potential biological functions of these BTF multi-omics features were explored using Gene set enrichment analysis (GSEA). A deep learning (DL)-based imaging segmentation model was built to extract multi-modal MRI radiomic features for 61 of the BC patients with MRI data. Two explainable tools (Gradient and Gradient*Input) were embedded into the DL model structure to explore biological implications of the radiomic features. Predictive least absolute shrinkage and selection operator (LASSO) models were trained to translate the radiomic features from the BTF multi-omics features for the BC patients without MRI data. Survival analyses were then performed to estimate the prognostic significance of each radiomic feature. Statistical mediation analyses were performed to further explore the underlying biological mechanisms of the identified biomarkers. Traditional semi-auto radiomic features and previously established single-omics features (e.g., BC risk gene expressions, pathway activity scores, and gene signature scores calculated from the gene expression profile) were used as baselines for comparison.
Results: Saliency maps of the multi-modal MRI radiomic features showed the excellent explainability of the built DL models. Three DL-based multi-modal MRI radiogenomic biomarkers were successfully identified. which were confirmed to have significant differences in overall survival (log-rank test, Bonferroni corrected *P* valueAPITD1, *HNF4*) and several metabolism related pathways (Purine metabolism pathway and Tryptophan metabolism pathway), which has a significant mediation effect on the relationship between one specific BTF multi-omics feature, representing the function of natural killer cells based on the GSEA analysis, and the BC survival time (adjusted *P* value**Conclusion:** The results may promote MRI as a non-invasive examination for BC prognosis and multi-level molecular status, and ultimately increase precision in BC prognosis and improve patient care.

PrgmNr 1484 - Identity-by-descent mapping in biobank-scale datasets

View session detail

Author Block: **L. E. Petty**[1], R. Bohlender[2], H.-H. Chen[3], J. Baker[4], G. Evans[4], J. E. Below[1], C. D. Huff[5]; [1]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, [2]1155 Pressler Street, Houston, TX, [3]Vanderbilt University Medical Center, Nashville, TN, [4]Nashville, TN, [5]U.T. MD Anderson Cancer Ctr, Houston, TX

Disclosure Block: **L.E. Petty:** None.

Identity-by-descent (IBD) mapping is a gene mapping approach that utilizes IBD segments to identify loci that are enriched for sharing in disease cases compared to controls. This approach allows for detection of loci that harbor variants that are too low frequency for imputation to perform well, which may have a greater penetrance than common variants well-powered for genome-wide association studies (GWAS).

We have developed a tool for performing biobank-scale IBD mapping, IBDMap. We leverage IBD segment data, detected using an external segment detection tool, and perform permutation-based testing to determine enrichment of IBD sharing by comparing rates of sharing in case-case pairs to rates in case-control pairs at each segment breakpoint, genome wide. Calculations for each breakpoint are performed independently and parallelized via a process queue. A complementary Python package performs map-reduce on multiple rounds of permutations, for computational efficiency in large-scale datasets. IBDMap performs flexible multiple testing correction, utilizing a permutation-based family-wise error or false discovery rate approach to account for the correlation structure. We also introduce a new haplotype-based approach which filters IBD segments that are unlikely to harbor rare, pathogenic variants due to high population frequency.

We applied IBDMap to a range of cardiovascular phenotypes in BioVU, Vanderbiltâ□□s DNA biobank with linked electronic health records. Using diagnostic billing codes, we determined case status for 69,819 European ancestry individuals genotyped on the MEGA^{EX} array for myocardial infarction, atherosclerosis, dyslipidemia, dilated cardiomyopathy, and atrial fibrillation. We then applied IBDMap for each phenotype to consensus IBD segments detected using iLASH, hap-ibd, and GERMLINE. We identified regions on chromosome 2 and 8 with genome-wide significant enrichment of IBD sharing in atherosclerosis cases.

In a â□□gold standardâ□□ set of 112 genes with 207 established, highly penetrant pathogenic variants for Mendelian diseases, all but one were detected by applying IBDMap to the traits in BioVU. In GWAS of the traits in the same samples using Firth regression in PLINK2, only 83% of the genes had at least one variant significantly associated with the trait, demonstrating greater power for detection of these highly penetrant loci in IBD mapping. We evaluated the IBDMap haplotype filtering approach with these 112 genes, and found that it resulted in an ~2-fold increase in -log10 p-value in the same BioVU samples. Our results demonstrate the potential of our IBD mapping approach for gene discovery.

PrgmNr 1485 - Incorporating family disease history and controlling case-control imbalance for population based genetic association studies

View session detail

Author Block: **Y. Zhuang**[1], B. N. Wolford[2], K. Nam[3], W. Bi[1,4], W. Zhou[5], C. J. Willer[2], B. Mukherjee[1], S. Lee[1,3]; [1]Biostatistics, University of Michigan, Ann Arbor, MI, [2]University of Michigan, Ann Arbor, MI, [3]Seoul National University, Seoul, Korea, Republic of, [4]Department of Medical Genetics, School of Basic Medical Sciences, Peking University, Beijing, China, [5]Massachusetts general hospital, Broad Institute, Boston, MA

Disclosure Block: **Y. Zhuang:** None.

In the genome-wide association analysis of population-based biobanks, the low prevalence in most diseases results in low detection power. If a family disease history is collected, the power can be improved by inferring the disease risks in control samples, which will be used as phenotypes in association analysis. In the presence of related samples, however, the existing methods, such as LT-FH, fail to address increased phenotypic correlation among closely related samples due to similar family history. In addition, existing approaches cannot adjust for the unbalanced phenotypic distribution. We propose a new method, TAPE (mixed-model-based Test with Adjusted Phenotype and Empirical saddlepoint approximation), which controls for increased phenotype correlation by introducing an additional variance component for closely related samples and accounts for case-control imbalance by using empirical saddlepoint approximation. We show through simulation studies that TAPE is computationally efficient and gains greater power than common GWAS without using family disease history (SAIGE) while controlling type I error. In power simulation, TAPE showed 21.0% increase in average chi-square statistics and 12.1% increase in causal SNP detection than SAIGE. While LTFH also had increased power over SAIGE, it suffered type I error inflation especially when analyzing related samples with low disease prevalence and MAF (118 times inflation at alpha=5E-8). We applied TAPE to 10 binary traits in UK Biobank among 408,898 white British samples and identified 659 genome-wide significant clumped variants, among which 127 were with MAF

PrgmNr 1486 - SUMMIT: An integrative approach for better transcriptomic data imputation improves causal gene identification

View session detail

Author Block: **Z. Zhang**[1], J. R. Bradley[1], L. Wu[2], C. Wu[1]; [1]Department of Statistics, Florida State University, Tallahassee, FL, [2]University of Hawaii Cancer Center, Honolulu, HI
Disclosure Block: **Z. Zhang:** None.
Transcriptome-wide association studies (TWASs), which integrate expression reference panels with genome-wide association study (GWAS) results to discover gene-trait associations, have deepened our understanding of genetic regulation in many complex traits. However, the number of analyzable genes and thus the power of TWAS is largely determined by the size of expression reference panels. One obvious but administratively onerous approach is to combine individual-level data from several consortia to increase the sample size. However, privacy concerns and sample consents often preclude access to individual-level genetic data, making a pooled individual-level expression panel unavailable. In this work, we introduce SUMMIT, a novel method that makes it possible to integrate a summary-level expression reference panel with a much larger sample size into GWAS to identify associated genes. In brief, we build gene expression prediction models in blood based on the summary data released by the eQTLGen consortium, which is to date the largest meta-analysis in 31,684 blood samples from 37 cohorts. Compared with benchmark methods, MR-JTI, TWAS-Fusion, PrediXcan, and UTMOST, SUMMIT built more gene expression prediction models (10,026 with R^2> 0.01) and achieved significantly higher prediction accuracy in different quantiles (p -9 by K-S test). To evaluate the performance of identifying significant associations, we applied SUMMIT to the summary statistics from 25 GWAS. Compared with benchmark methods, SUMMIT identified substantially more associations for each trait analyzed, showing 222% improvement compared with MR-JTI ($p=3.1*10^{-9}$ by the Wilcoxon rank test), 306% improvement compared with TWAS-fusion, 264% improvement compared with PrediXcan, and 211% improvement compared with UTMOST. Next, we compared different methods in terms of identifying the likely causal genes that mediate the associations between GWAS loci and the traits of interest by using a set of 1,424 likely causal gene-trait pairs curated by using the OMIM. We show that the SUMMIT yielded good sensitivity and specificity for identifying the silver standard genes and achieved the highest AUC (0.701) among all the methods compared (MR-JTI 0.6329; PrediXcan 0.6164; TWAS-fusion 0.6041; UTMOST 0.6213). More importantly, SUMMIT was applicable to analyze genes with smaller heritable expressions (0.005 2

PrgmNr 1487 - Allele-specific expression of SNPs involved in the immune response and gene transcription is associated with BMI

View session detail

Author Block: **A. Keshawarz**[1], T. Huan[2], R. Joehanes[1], C. Yao[3], C.-T. Liu[4], C. Demirkale[5], V. Ramachandran[6], L. A. Cupples[7], D. Levy[8]; [1]National Heart, Lung, and Blood Institute, Framingham, MA, [2]University of Massachusetts Medical School, Worcester, MA, [3]Framingham Heart Study, Framingham, MA, [4]Boston University SPH, Boston, MA, [5]National Institutes of Health, Bethesda, MD, [6]Boston University School of Public Health, Boston, MA, [7]Boston Univ Sch Pub Hlth, Boston, MA, [8]NHLBI/NIH, Framingham, MA

Disclosure Block: **A. Keshawarz:** None.

Background. Body mass index (BMI) is an estimate of general adiposity associated with numerous clinical outcomes, including subclinical and clinical cardiovascular and metabolic disease. Growing genome-wide association studies have identified hundreds of BMI-associated loci, but the transcriptomic signature of BMI is incompletely understood. The objective of this study was to quantify the association between allelic imbalance and BMI using allele-specific expression analysis in conjunction with GWAS of BMI. **Methods.** Whole genome sequencing and RNA sequencing data were collected from 720 Framingham Heart Study (FHS) Offspring participants (59% women, mean age 66Â±8 years) and 954 FHS Third Generation participants (52% women, mean age 46Â±9 years) as part of the Trans-Omics for Precision Medicine (TOPMed) program. Heterozygous SNPs (n=780,599) were evaluated for allele-specific expression, and the ratio of reference allele to total allele counts for SNPs was calculated. SNPs with significant allelic imbalance based on a Bonferroni-corrected p-value (0.05/780,599) in a binomial test were subsequently evaluated in multiple linear regression to test the association between the reference allele/total allele count ratio and BMI after adjustment for age, sex, and family structure. SNPs significantly associated with BMI at FDR cis-expression quantitative trait loci (eQTLs) to investigate other known functions and associations. **Results.** After adjustment for age, sex, and family structure, allelic imbalance in 139 SNPs was significantly associated with BMI. SNPs identified were annotated to genes that were enriched for cell activation and immune response pathways, and the top five most significant SNPs identified were annotated to *FLYWCH1, ZDHHC6, MYADM*, and *TCL1A*. Twenty-seven SNPs were in regulatory regions; of these, 18 were in promoter or promoter flanking regions. Seven SNPs identified in these analyses were associated with traits in the GWAS catalog, including BMI. Furthermore, these SNPs overlapped with 181 unique *cis*-eQTLs (pPLGLB2, HLA-J, TRIM31, DHDDS, AL671883.2, DPY19L1P2, AL022345.4). **Conclusion.** SNPs implicated in the immune response pathway and in gene regulation show significant allelic imbalance associated with BMI; thus, allele-specific expression may partially explain population-level variation in BMI.

PrgmNr 1488 - Genetic determinants of prostate-specific antigen levels improve cancer screening utility

View session detail

Author Block: **L. Kachuri**[1], R. E. Graff[1], J. P. Shelley[2], K. Schaffer[2], J. D. Mosley[2], P. J. Goodman[3], C. Till[3], I. Thompson[4], R. Klein[5], S. K. Van Den Eeden[6], T. J. Hoffmann[1], J. S. Witte[7]; [1]University of California San Francisco, San Francisco, CA, [2]Vanderbilt University Medical Center, Nashville, TN, [3]Fred Hutchinson Cancer Research Center, Seattle, WA, [4]Christus Health, San Antonio, TX, [5]Icahn School of Medicine at Mt. Sinai, New York, NY, [6]Kaiser Permanente Northern California, Oakland, CA, [7]Univ California San Francisco, San Francisco, CA

Disclosure Block: **L. Kachuri:** None.

Prostate-specific antigen (PSA) testing is controversial due to issues related to sensitivity and specificity, resulting in overdiagnosis and overtreatment of prostate cancer (PCa). Genetic determinants of PSA in cancer-free men could be used to correct observed PSA values by accounting for PSA variation that does not reflect PCa.

We conducted the largest ever genome-wide association study (GWAS) of PSA in men without PCa (N=65,962; 63,338 European ancestry) using longitudinal measures from the UK Biobank (UKB; n=26,491), BioVU (n=8078), and Genetic Epidemiology Research on Adult Health and Aging (GERA; N=30,088) cohorts. Our GWAS discovered 87 variants associated with PSA levels (P-8), including 37 loci independent of previously reported associations (r^2PSA) constructed from these variants was validated in the Prostate Cancer Prevention Trial (PCPT; P=8.7Ã□10^{-54}). In the PCPT, which enrolled PCa-free men with PSAâ□¤3 ng/mL, PRS explained a larger proportion of PSA variation than age (4.0% vs. 1.3%).

Consistent with the hypothesis that genetic predisposition to elevated PSA increases the detection of low-grade PCa, PRS_{PSA} was inversely associated with Gleason score (â□¤6 vs. â□¥8: OR=0.83, P=8.4Ã□10^{-5}; 6485 GERA cases) and PCa mortality (HR=0.82, P=7.4Ã□10^{-9}; 8834 UKB cases). We then evaluated how genetically corrected PSA values affect reclassification at cut-offs used for biopsy recommendations in GERA. Among non-cases with a negative biopsy, 18.4% were reclassified below the referral threshold, while 2.7% moved upward. In cases, downward reclassification (3.9%) was higher than upward (1.9%) when considering PSA values â□¥2 years before diagnosis. This trend was more pronounced in cases with low-risk disease (Gleason Lastly, we explored the role of PSA-related selection bias on associations with PCa risk and mortality. There is substantial sharing of genetic loci between PSA and PCa, illustrated by the high correlation between their genetic scores (r=0.287, P-500) in UKB (n=164,669 not included in the PSA GWAS). Although this may partly reflect pleiotropy, there was evidence of bias due to screening (P=2.1Ã□10^{-130}). Re-fitting PRS_{PCa} using bias-corrected risk allele weights attenuated its correlation with PRS_{PSA} (r=0.049, P=5.5Ã□10^{-93}) and revealed a previously absent association with PCa mortality (HR=1.18, P=0.035).

Our work provides preliminary evidence that genetic correction of PSA levels may improve PCa screening. Larger and more diverse study populations are required to fully characterize the genetic basis of PSA variation and optimize its clinical utility.

PrgmNr 1489 - Joint intron splicing-based transcriptome-wide association study identifies new candidate susceptibility genes for breast cancer

View session detail

Author Block: **G. Gao**[1], P. Fiorica[1], A. N. Barbeira[2], J. McClellan[1], D. Huo[1], H. Im[2]; [1]Department of Public Health Sciences, University of Chicago, Chicago, IL, [2]The University of Chicago, Chicago, IL
Disclosure Block: **G. Gao:** None.
In this study, we proposed a joint intron splicing-based transcriptome-wide association study approach (IntronXcan) that combined information from multiple excised introns in a gene across multiple tissues. IntronXcan used splicing prediction models trained in 47 tissues in the GTEx (v8) data with a multivariate adaptive shrinkage (mash) method, which can jointly estimate effects of splicing quantitative trait loci (sQTLs) in multiple tissues, accounting for correlation among nonzero sQTL effects in different tissues. We applied IntronXcan to the GWAS summary statistics from the Breast Cancer Association Consortium (BCAC) of about 229,000 women of European ancestry. We identified 550 genes significantly (P -^{6}) associated with breast cancer. We replicated 75 of these 550 genes at significance level (P -^{5}) when applied IntronXcan to GWAS summary statistics from an independent dataset on breast cancer extracted from UK Biobank. To determine if these genes are independent of previously identified breast cancer loci, we performed a conditional and joint analysis (COJO) by conditioning summary statistics on previously published (risk) index SNPs within +/- 10 Mb of each significant gene. We then re-performed IntronXcan using the conditional summary statistics. Results showed that 31 of the 75 replicated genes were still significant. We treat these 31 genes as conditionally independent of known breast cancer GWAS loci. Of these 31 genes, 13 novel candidate susceptibility genes are located beyond 500kb from their nearest GWAS index SNP. We similarly performed gene expression-based joint TWAS (S-MultiXcan) that integrated predicted expression data from multiple tissues. We identified 311 genes significantly associated with breast cancer using BCAC summary statistics (P â□□6). Among them, 23 genes were replicated (P â□□5) in our S-MultiXcan analysis of the independent UKB dataset. Six of the 23 replicated genes remained significant after conditioning on previously identified index SNPs. We further performed a meta-analysis to combine BCAC and UKB GWAS summary statistics and then applied splicing-based IntronXcan and expression based S-MultiXcan to the meta-analysis summary statistics. We identified an additional 162 and 68 significant genes, respectively. Our analyses indicated that the splicing based IntronXcan could identify a greater number of significant genes than the gene expression-based S-MultiXcan and individual intron splicing-based TWAS.

Poster Presentations

PrgmNr 2001 - ANGPT1 TEK signaling pathway and primary congenital glaucoma A likely impact in a female newborn glaucoma associated to a dup8q22

View session detail

Author Block: **B. Abdelmoula**, N. Bouayed Abdelmoula; Genomics of Signalopathies at the Service of Med., Med. Univ., Sfax, Tunisia, Sfax, Tunisia
Disclosure Block: **B. Abdelmoula:** None.
The elevated pressure in primary congenital glaucoma occurs subsequently to defects in the aqueous humor outflow passageway. Secretion of ocular anterior chamber fluid produced by the ciliary body is exhausted primarily through Schlemmâ□□s canal and uveoscleral passages. Here, we describe a neonate female conceived from a consanguineous couple who lost two early pregnancies and two males suffering from severe immunologic disorders, at the age of 2 and 7 months. She harbored severe facial dysmorphism, multiple malformations, postnatal growth impairment and a delayed psychomotor development with severe hypotonia and seizures crisis. She had in particular multiple ocular defects with hypertelorism, buphthalmos as in congenital glaucoma (diffuse enlargement of the eye with exophthalmia and a large cornea with a grey corneal arcus) with blindness related to optic nerve disorder. She had furthermore many angiomas and blue birthmarks related to venous dilatations, which were disseminated through the majority of her body skin. Standard karyotyping showed an unbalanced translocation with a deletion of the critical region of Wolf-Hirschhorn syndrome within 4p15.3-pter and a duplication of the 8q22-23 chromosomal region, subsequently to a transmitted der(4)t(4;8)(p15.3;q22-23) from her mother. Recently, ANGPT1-TEK signaling is demonstrated to be required for Schlemmâ□□s canal development in mice. Furthermore, hemizygosity for Tek in mice, led to the formation of severely hypomorphic Schlemmâ□□s canal and trabecular meshwork, as well as elevated intraocular pressure, demonstrating that anterior chamber vascular development is sensitive to Tek gene dosage and the resulting decrease in ANGPT1-TEK signaling. Even as ANGPT1 is the primary ligand of TEK in the irido-corneal angle and the inactivation or perturbation of ANGPT/TEK pathway leads to degeneration and/or irregular canal formation, increased intraocular pressure and glaucoma during development, abnormal ANGPT1 gene dosage and the resulting effect in ANGPT1-TEK signaling pathway may be expected as an alternative mechanism of neonate congenital glaucoma. In our case report, this abnormal gene dosage following ANGPT1 duplication may resulting in loss-of-function defects leading to glaucoma as well as venous malformations. To note that, gain-of-function TEK mutations were also linked to hereditary and sporadic venous malformations.

PrgmNr 2002 - Chromosomal abnormality -17/17P- in acute B-lymphoblastic leukemia (B-ALL) implies loss of heterozygosity of *TP53* gene

View session detail

Author Block: **J. Yan**, Q. Zheng, Y. Zhang, N. Chen, Z. Yin, X. Jia, D. Wang; Beijing Boren Hosp., Beijing, China
Disclosure Block: **J. Yan:** None.
Loss of heterozygosity (LOH) of tumor suppressive genes is an important factor for poor curative effect and prognosis in hematologic malignant diseases.The purpose of this research was to study the cases of B-ALL with *KMT2A-AFF1* fusion accompanied by deletion of chromosome 17 or the short arm of chromosome 17 (-17/17P-) and *TP53* gene mutation, and to preliminarily evaluate the occurrence of LOH and its impact on the outcome of these patients. The data of B-ALL patients with chromosomal 4 and 11 translocation [t(4;11)(q21;q23)] admitted to Beijing Boren Hospital from January 2019 to March 2021 were collected and analyzed. Additional abnormality of deletion of the *TP53* locus detected by FISH and *TP53* gene mutation detected by next-generation sequencing (NGS) together with clinical information was evaluated. Twenty-five patients with t(4;11)(q21;q23) of B-ALL were investigated. Fourteen of these patients are males and 11 are females, and their ages are from 5 months to 45 years (median age 10 years). Among the 25 cases, 23 cases were examined by FISH technique using P53 probe, and the deletion of *TP53* locus (positive) were found in 13 cases (56.5%). *TP53* gene mutation was detected in 13 of the 17 investigated cases (76.5%) by means of NGS technology. All the 25 patients received multicourse treatment. Among them, only 2 were first-diagnosed B-ALL, while all the rest had relapse in different clinical course. Seven of the 25 patients had once achieved temporary remission under different treatment regiments, and 4 of the 7 patients did not carry *TP53* gene mutation nor -17/17p- as indicated by sequencing technology and cytogenetics, respectively. This retrospective study demonstrated that the t(4;11)(q21;q23)/*KMT2A-AFF1* fusion gene-bearing B-ALL patients had frequent additional chromosomal abnormalities, of which the most common one was -17/17P-. Deletion of *TP53* loci detected by FISH in the same group of patients was consistent with karyotype results. Pathogenic *TP53* mutations, which was highly frequent among these patients as indicated by NGS was also highly coincidence with the FISH results. These findings suggest that the patients lost normal anticancer function of *TP53* gene due to LOH. Considering the high relapse frequency in this B-ALL group, we propose that additional genetic abnormalities, especially -17/17P-, tend to evolve during the disease development in t(4;11)(q21;q23)-bearing B-ALL patients, resulting in the LOH of *TP53* gene, which is directly related to the poor prognosis of this type of leukemia. Therefore, the deletion of *TP53* locus on chromosome 17 demonstrated by cytogenetic analysis implies LOH in the B-ALL.

PrgmNr 2003 - Clinical features and genetic analysis of pediatric pilomixoid astrocytomas

View session detail

Author Block: **D. Colak**[1], E. AlShail[1], M. Alsagob[2], A. N. Alahmari[1], H. Al-Hindi[1], A. A. M. Dababo[1], Z. Al Masseri[1], R. AlSalamah[1], E. Almohseny[1], A. Alduhaish[3], A. Ul-Haq[1], H. Khalil[1], N. Kaya[1]; [1]King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia, [2]KACST-BWH/Harvard Ctr. of Excellence for Biomedicine, Joint Ctr.s of Excellence Program, King Abdulaziz City for Sci. and Technology, Riyadh, Saudi Arabia, [3]AlFaisal Univ., Riyadh, Saudi Arabia

Disclosure Block: **D. Colak:** None.

Pilocytic astrocytomas (PAs) are the most frequently encountered brain tumors with peak incidence in the first decade of life. Recently, a distinct group of tumors previously diagnosed as PA were identified and defined as pilomyxoid astrocytoma (PMAs) that presents unique histological characteristics and have more aggressive clinical course. The studies on genetics of PMA are scarce. In this study, we focused on one of the largest cohort of PMA samples to date to understand the genetics and cytogenetic makeup of this solid type of tumor. We examined and compared the clinical outcome of the 27 pediatric patients who were diagnosed with PA and PMA. These samples were also cytogenetically tested using Affymetrixâ□□s oncoscan arrays. The PMAs share alike features with PA; however, they were more aggressive and had shorter overall survival rate than the PAs. Our study yielded a previously reported duplication (KIAA1549-BRAF Fusion) in most of the tested samples. Besides the fusion gene, 11 samples had additional cytogenetic abnormalities. In conclusion, this study is the first report of Saudi patients with PMA/PA, and provides detail clinical picture and outcome of these rare pediatric tumors in Saudi population and hence may help better diagnosis and characterization of PMA.

PrgmNr 2004 - CYP polymorphisms and heart failure

View session detail

Author Block: **N. B. Abdelmoula**[1], R. Abdelhedi[2], UR17ES36 Team; [1]Genomics of signalopathies at the service of medicine UR17ES36, Med. Univ. of Sfax, Sfax, Tunisia, Sfax, Tunisia, [2]Laboratoire de procÃ©dÃ©s de criblage molÃ©culaire et cellulaire LR15CBS07, Ctr. of Biotechnology of Sfax, Sfax, Tunisia, Sfax, Tunisia

Disclosure Block: **N.B. Abdelmoula:** None.

Coronary artery disease is a complex multifactorial disorder. Its pathophysiology depends on a large number of lifestyle and environmental risk factors. Numerous genes and single nucleotide polymorphisms were reported to be associated, such as VEGF, VEGFA, SCARB1 and CYPs. Despite advances in stent technology, patients with coronary artery disease remain at risk for cardiovascular mortality, mainly because of myocardial infarction and development of heart failure. The aim of this study was to assess the impact of the loss-of-function polymorphisms of CYP2C19 in one hundred Tunisian patients affected by coronary artery disease. We enrolled 100 patients with acute coronary syndrome under-going percutaneous coronary stenting. Genotypes were determined using PCR-RFLP techniques. SNP genotyping of CYP2C19*2 (rs4244285) and CYP2C19*3 (rs4986893) was performed. PCR amplification was carried out using two couples of primers and RFLP was carried out using respectively Sma and BamHI restrictions enzymes. Data concerning demographic parameters, coronary heart disease risk factors and treatment were collected for each patient. The occurrence of major cardiovascular events between in the carriers and the non-carriers of SNPs was recorded. Comparison of categorical variables was performed using the chi-square test. Values were considered statistically significant at p A showed that 77% of patients were homozygous G/G and 23% were heterozygous G/A. The 636G>A polymorphism was absent in our population. After percutaneous coronary intervention, 3,9% of G/G patients versus 4,3% of G/A patients developed a left ventricular failure. All of them were females but without any statistical significant difference. When all major adverse cardiac events post-percutaneous coronary intervention are considered, a statistical significant difference (p=0.075) among the G/G and G/A groups was recorded. CYP2C19*2 carrier status associated with an increased risk of heart failure following coronary stent placement is not well reported. The association between CYP2C19*2 polymorphism and inflammatory markers concentrations could suggest that CYPs including CYP2C19 seem to be candidate genes for cardiovascular risks. It was shown recently that CYP2C19 genotype may be a risk factor for coronary microvascular disorder via inflammation but exclusively in the female population. A larger study is needed to better assess the role of CYP2C19 genotyping as an independent risk factor in the development of heart failure following percutaneous coronary intervention via an inflammarory process.

PrgmNr 2005 - Geographic differences of the distribution of the mutational spectrum of *BRCA1* and *BRCA2* genes in Algerian population

View session detail

Author Block: **F. Cherbal**[1], C. Mehemmai[1], H. Gaceb[1], W. Benbrahim[2], H. Mahfouf[3], K. Boualga[4]; [1]Molecular Genetics Team, LMCB, Faculty of Biological Sci., USTHB, Algiers, Algeria, [2]Med. Oncology Service, Anti-cancer center of Batna, Batna, Algeria, [3]Mohamed El Kolli Publ. Hosp., Academic Med. Oncology Services, Sch. of Med., Univ. of Algiers1, Rouiba, Algeria, [4]Radiation Oncology Service and Lab. for Cancer Res., Anti-cancer Ctr. of Blida, Blida, Algeria

Disclosure Block: **F. Cherbal:** None.

Background: Algeria is a country of continental dimension with admixed population of Berber descent with Sub-Saharan African, European and Middle East elements. The knowledge of recurrent mutations in *BRCA* genes and their geographical distribution is useful for the design of an efficient and affordable genetic testing for hereditary breast and ovarian cancer (HBOC) patients and families at risk. In this retrospective study, we report the mutational spectrum of *BRCA* genes in HBOC patients from North central region and North East region of Algeria which is called ' The Aures region', respectively. **Methods**: first**,** we analyzed the complete sequence of *BRCA1* and *BRCA2* of 70 HBOC families from North central region of Algeria using HRM-Sanger sequencing and MLPA techniques. Then, *BRCA* genes were screened using PCR-Sanger sequencing in a second cohort of 56 HBOC patients from the North central region and a cohort of 113 HBOC patients (109 women and 4 men) from the Aures region, respectively. The screening included all exons where a recurrent mutation was previously found in BRCA1 exons (3, 4 and 10) and in *BRCA2* e x o n (1 0) in our first screening. In addition, twelve HBOC patients from North East region were analyzed using a NGS analysis with a cancer panel of 30 hereditary cancer genes or BRCA1/2 genetic test. **Results:** We detected 4 recurrent mutations in the *BRCA*1 gene of HBOC patients from the North Central region: c.83_84delTG (7 unrelated families), c.181T>G (3 unrelated families), c.798_799delTT (3 unrelated families), c.2125_2126insA (3 unrelated families), Del exon 7 (2 unrelated families) and one recurrent mutation in the *BRCA2* gene : c.1310_1313delAAGA (2 unrelated families), respectively. The recurrent mutations in *BRCA1* and *BRCA2* genes already reported in HBOC patients from North central region of Algeria have not been detected in patients from the North East region. In addition, NGS analysis of *BRCA* genes in HBOC patients from North East region revealed 2 distinct mutations in *BRCA1* gene: Del exon 15 (2 unrelated families), c.5332 + 1G > A (one family) and 4 distinct mutations in *BRCA2* gene: c.1813dupA (one family), c.7654dupA (one family) c.8485C>T(one family) and c.8940delA (one family). Interestingly, all these mutations detected in *BRCA* genes in HBOC patients from the Aures region have never been reported in Algerian population. **Conclusions**: Our results showed differences in the distribution of the mutation spectrum of BRCA genes between the North central region and the Aures region of Algeria. The present study will help to implement affordable genetic testing and to improve the clinical management and better risk assessment of hereditary breast and ovarian cancer.

PrgmNr 2006 - Germline mitochondrial DNA non-coding region variants among sporadic breast cancer patients of two Sri Lankan ethnic groups

View session detail

Author Block: **J. T. Kotelawala**[1], K. H. Tennekoon[1], R. A. C. Ranasinghe[1], H. A. C. I. K. Rodrigo[1], G. K. S. De Silva[2], W. A. H. A. Perera[3], N. Yoganathan[3], M. R. S. Manatunga[3], D. M. A. S. Dissanayake[4], N. Joseph[4,5], C. Rajasooriyar[4,5], K. Indranath[4,5]; [1]Inst. of Biochemistry, Molecular Biology and Biotechnology. Univ. of Colombo, No. 90. Cumaratunga Munidasa Mawatha, Colombo 03, Sri Lanka, [2]Natl. Cancer Inst., Maharagama, Sri Lanka, [3]Kandy Teaching Hosp., Kandy, Sri Lanka, [4]Jaffna Teaching Hosp., Jaffna, Sri Lanka, [5]Tellippalai Base Hosp., Jaffna, Sri Lanka

Disclosure Block: **J.T. Kotelawala:** None.

Breast cancer is the main cancer occurring amongst women in Sri Lanka accounting for nearly 25% of cancers diagnosed. Studies have shown association of mitochondrial (mtDNA) variations with cancer. We previously reported mitochondrial variants among breast cancer patients and controls of Sri Lankan Sinhalese ethnicity, however these variants failed to show a significant association with breast cancer within the study sample. Interestingly, initial studies revealed that some variations such as 16189C occurred more frequently in patients of Sri Lankan Tamil ethnicity compared to Sinhalese and Sri Lankan Moor ethnicities, hence a larger sample was studied to determine if these variations are significantly associated with breast cancer. Here we report the mtDNA noncoding variants in 30 sporadic breast cancer patients and healthy control pairs each of Sri Lankan Tamil and Sri Lankan Moor ethnicities.

Genomic DNA was extracted from peripheral venous blood and PCR amplified using two pairs of mtDNA non-coding region specific primers. The amplified PCR product was sequenced using Sanger sequencing. The sequences obtained were compared against the revised Cambridge Reference Sequence (NC_012920.01) using Mutation Surveyor (v.4.0).

The following variants were commonly observed in patients and controls of Sri Lankan Tamil (SLT) ethnicity 73G (29 vs 28); 146C (8 vs 2); 152C (8 vs 10); 263G (28 vs 30); 489C (21 vs 13); 523-524 deletion (6 vs 5); 523 AC insertion (1 vs 1); 16189C (7 vs 2); 16223T (21 vs 6); 16311C (7 vs 4); 16519C (24 vs 24). The following variants were commonly observed in patients and controls of Sri Lankan Moor (SLM) ethnicity: 73G (24 vs 27); 146C (5 vs 8); 150T (7 vs 7); 152C (12 vs 10); 263G (30 vs 30); 489C (15 vs 11); 523-524 deletion (7 vs 4); 523 AC insertion (0 vs 2) 16051G (7 vs 9); 16189C (3 vs 1); (16223T (17 vs 14); 16311C (10 vs 5); 16352C (7 vs 4); 16497G (4 vs 6); 16519C (18 vs 15). Some variations occurred more frequently in patients. The prevalence of mtDNA variant 16223T was significantly higher in the patients than in the controls (Fischerâ□□s exact test: p=0.0002) among Sri Lankan Tamils but not among Sri Lankan Moors (present study) or Sinhalese (as previously reported). Although mtDNA variant 16223T can be suggested as a predictive biomarker for sporadic breast cancer among Sri Lankan Tamils, a larger sample is needed to confirm these findings before its clinical use. This study also highlights the importance of studying individual ethnic groups prior to recommending potential biomarkers in disease prediction.

This research is supported by the National Science Foundation of Sri Lanka (NSF/SCH/2016/04)

PrgmNr 2007 - Identification of RP gene and ERÎ± gene polymorphism as genetic markers for the improvement of breast cancer management in cameroon

View session detail

Author Block: **N. NGUEDIA KAZE**; Univ. of Yaounde I, YAOUNDE, Cameroon
Disclosure Block: **N. Nguedia kaze:** None.
Background : Breast cancer is a real public health problem in Cameroon, where more patients with this cancer usually die a year after diagnosis, as it is still based on histological examination, mortality due to that cancer is far from decreasing. **Rational**: Since cancer is an accumulation of molecular changes in gene, the +331 GA polymorphism of PR gene and 397TC and 351AG polymorphisms of ERÎ± gene could be investigated as breast cancer etiological factors. **Methods :**We carried out a case control study, in which 16 cases diagnosed positive for breast cancer at the Yaounde general hospital was recruited by consulting the patient file and Blood samples was also collected from them and 22 healthy women who were recruited using a questionnaire and an inform consent signed each of them. The +331 GA polymorphism in the PR gene was identified using NIaIV endonuclease and by PCR-RFLP and direct molecular haplotyping was used to determine the relationship between the 2 polymorphisms in the ERÎ± gene, 397TC and 351AG by restriction fragment-length polymorphisms. The data were analyzed using Statistical Package Social Sciences v20. **Results**: The population was mostly postmenopausal with an invasive ductal carcinoma with an average age of 53+-11,08 years. A mutant AA genotypic frequency of the PR gene was high in both, the case and control groups 100% and 94% respectively. We obtained an Odds Ratio of 1.067 with 95% Confidence Interval of 0.940-1.211 and P value of 0.421, making a non-significant association with the risk of developing breast cancer. In addition, we identified the mutant CG haplotype allele of the ER alpha gene with a predominance in the cancer group exclusively with a frequency of 56.25% with an OR of 2.022 with a 95% CI 0.432 - 9.461 with a p-value of 0,3, not significantly associated with breast cancer. However, the CG.CG mutant genotype of the ER alpha gene (31.25%) in the diseased population was significantly associated with breast cancer with an OR 1.455, CI 1.045-2.024 and a p-value of 0.05. **Conclusion**: This shows that the polymorphisms studied would predispose to a risk of developing breast cancer.

PrgmNr 2008 - Low Prevalence of Large Genomic Deletions and Duplications Observed in the Sequential Use of MLPA after Negative NGS Result in Thai Breast Cancer Cohort

View session detail

Author Block: **P. Mutirangura**[1], P. Lertwilaiwittaya[2], E. Routhumnong[1], S. Wiboonthanasarn[1], K. Pongsuktavorn[1], P. Dungort[1], C. Limwongse[3], W. Thongnoppakhun[1], M. Pithukpakorn[1]; [1]Faculty of Med. Siriraj Hosp., Bangkok, Thailand, [2]Mahidol Univ., Bangkok, Thailand, [3]Bangkok, Thailand
Disclosure Block: **P. Mutirangura:** None.
Introduction*BRCA1* and *BRCA2* mutations are the most common cause of hereditary breast cancer. Next generation sequencing (NGS) has been adopted to screen these genes in clinical scenarios. It is known that single nucleotide variants (SNVs) and small insertions/deletions (indels) contribute to the majority of all *BRCA1/2* mutations. However, large genomic deletions/duplications (del/dup), accounted for 10 percent of overall mutation mechanism in Western population, might be missed from the use of such technique. We aimed to report the diagnostic yield of the use of sequential multiplex ligation-dependent probe amplification (MLPA) to detect large genomic del/dup in *BRCA1/2* in Asian population.
MethodsA retrospective review of 440 breast cancer patients underwent NGS at Siriraj Hospital in 2016-2021 was done. Information including demographic data, patient characteristics, type of breast cancer, and result of genetic identification were recorded. The patients with high clinical suspicion or positive JSI yet negative NGS test underwent MLPA test. The patients with high clinical or bioinformatics (suggested by JSI medical systems or demonstrate a significant drop in sequencing read depth compared to pooled sample) suspicion were selected and sent for MLPA. We use descriptive statistics to analyze the yield of each criterion.
ResultsAmong 440 patients in this study undergoing NGS test, 105 patients (23.8%) had positive results of pathogenic or likely pathogenic variant in *BRCA1/2* genes.There were 66 patients with negative NGS underwent MLPA. Thirty patients (44%) had bioinformatics suspicion, while 36 patients (56%) had high clinical suspicion. Only one patient (1/66, 1.5%) had a positive result of *BRCA1* mutation in clinical suspicion group.
Discussion/ConclusionNGS demonstrated a high yield of SNVs and small indels in Thai populations (105/440 patients; 24%). Large genomic del/dup contribute to only a small percentage (1/66 patients; 1.5%) when compared to the Western populations. The sequential use of MLPA after a negative NGS result may not be fruitful in Asian populations. Further study should be done to explore other screening criteria.

PrgmNr 2009 - Renal cancer progression & contribution of chromatin remodeling complexes

View session detail

Author Block: **W. SMAOUI**[1,2], N. BOUAYED ABDELMOULA[3,2]; [1]Habib BOURGUIBA Univ. Hosp., Sfax Univ., TUNISIA, SFAX, Tunisia, [2]UR17ES36, Sfax, Tunisia, [3]Faculty of medecine of sfax, Sfax Univ., TUNISIA, SFAX, Tunisia
Disclosure Block: **W. Smaoui:** None.
Background:Renal cell carcinomas accounts for around 95% of the neoplasms arising from the kidney. These malignant adenocarcinomas initiates in the renal tubular epithelium and progresses leading to death of 40% of patients. Different histological subtypes are described including clear cell, papillary, and chromophobe renal cell carcinomas. Loss-of-function mutations in chromatin modifiers are now recognized as common in renal cell carcinomas and some of them play a central role in tumor progression. Our ongoing study includes 180 localized renal tumor samples from Tunisian patients, for which screening of recurrent and hot spot mutations of PBRM1, BAP1, SETD2, KDM5C and KMD6A genes will be conducted. Here, we focused our attention on the literature background of recurrent mutations of SETD2.Material and methods:A review of the literature has been conducted to state recurrent mutations of SETD2 in RCC as well as molecular techniques that can be applied according to our technical platform aptitudes.Results:We found that DNA sequencing revealed that mutations in SETD2 occur in 3 to 12 percent of clear-cell renal cell carcinomas cases (the third-most-commonly mutated gene) and are associated with poor clinical outcome. PCR primer sets to amplify and sequence all coding sequences and exon-intron boundaries of SETD2 as well as PCR amplification conditions and sequencing protocols using Sanger methodology are available. Recurrent pathogenic mutations in renal cell carcinomas samples have been reported.Conclusion:SETD2 functions as a histone methylation and as a microtubule methyltransferases. SETD2 is coding for a multi-domain protein with a central SET domain responsible of the methyltransferase activity and a C-terminus SRI domain mediating the interaction between SETD2 and the phosphorylated C-terminal domain of RNA polymerase II. Hundreds of distinct SETD2 mutations have been identified across a wide range of human tumors, including epithelial, nervous and hematopoietic types. SETD2 mutation was first described in clear cell renal cell carcinomas and the most common and recurrent mutations are described in SET domain suggesting loss of catalytic activity such as the R1625C and in SRI domain such as the R2510H.

PrgmNr 2010 - Small nucleolar RNAs as potential predictive/prognostic markers for breast cancer

View session detail

Author Block: **J. M. Hartikainen**[1,2], E. KÃ¤rkkÃ¤inen[1,2], S. Heikkinen[3,4], M. TengstrÃ¶m[5,6], V-M. Kosma[1,2,7], A. Mannermaa[1,2,7]; [1]Sch. of Med., Inst. of Clinical Med., Pathology and Forensic Med., Univ. of Eastern Finland, Kuopio, Finland, [2]Translational Cancer Res. Area, Univ. of Eastern Finland, Kuopio, Finland, [3]Sch. of Med., Inst. of Clinical Med., Univ. of Eastern Finland, Kuopio, Finland, [4]Sch. of Med., Inst. of Biomedicine, Univ. of Eastern Finland, Kuopio, Finland, [5]Sch. of Med., Inst. of Clinical Med., Oncology, and Cancer Ctr. of Eastern Finland, Univ. of Eastern Finland, Kuopio, Finland, [6]Cancer Ctr., Kuopio Univ. Hosp., Kuopio, Finland, [7]Biobank of Eastern Finland, Kuopio Univ. Hosp., Kuopio, Finland
Disclosure Block: **J.M. Hartikainen:** None.
The small nucleolar RNAs (snoRNAs), including the small Cajal body-specific RNAs (scaRNAs), have recently been suggested as intriguing players in cancer. To study their role in breast cancer (BC), we performed small RNA-sequencing in 195 fresh-frozen invasive BC and 20 benign breast tissue samples from the Eastern Finnish Kuopio Breast Cancer Project (KBCP), to explore differentially expressed snoRNAs with potential prognostic and predictive value in invasive local BC. Total RNA was extracted from fresh-frozen tissues using the Ambion mirVana miRNA Isolation Kit, processed using the Illumina TruSeq Small RNA Library Prep kit and sequenced. Bioinformatic analysis of non-miRNA/piRNA sncRNA expression included read quality assessment (FastQC), adapter trimming (TRIMMOMATIC), removal of e.g. rRNA reads (Bowtie2), and alignment (Tophat2) to human reference sncRNA transcriptome (subset of GENCODE v22). Differentially expressed sncRNAs were identified using DESeq2 in R from gene-wise read counts. Survival was analysed using survival::coxph in R.
Utilizing the comprehensive background, clinical, treatment, and follow-up data extending up to 25 years in the KBCP samples, we identified altogether 58 snoRNAs and scaRNAs that were differentially expressed in invasive local BC compared to benign breast tissue (P_{Adj}PPPP$_{Adj}$

PrgmNr 2011 - Studying the phenotypic variability of ERCC6L2 deficiency in a genetic island

View session detail

Author Block: **L. Kalfon**[1], S. Ben Haroush[1], H. Tamary[2], A. Kuperman[3], A. Braester[4], A. Singer[5], T. C. Falik-Zaccai[1,6]; [1]Inst. of Human Genetics, Galilee Med. Ctr., Nahariya, Israel, [2]Rabin Med. Ctr., Petah Tikva, Israel, [3]Pediatric Hematology, Galilee Med. Ctr., Nahariya, Israel, [4]Hematology Inst., Galilee Med. Ctr., Nahariya, Israel, [5]Dept. of community genetics, Israeli Ministry of Hlth., Tel Aviv, Israel, [6]The Azrieli Faculty of Med., Bar Ilan, Safed, Israel
Disclosure Block: **L. Kalfon:** None.
ERCC6L2 belongs to the Snf2 family of helicase-like proteins that are involved in nucleotide excision repair, mitochondrial function and transcription regulation. Pathogenic variants in this gene have been related to bone marrow failure (BMF) with variable phenotypic expression, age of onset and somatic genetic findings in bone marrow. Up to date about 16 cases with bi-allelic variants in *ERCC6L2* have been reported in the literature. Clinical representation varied from myelodysplastic syndrome (MDS) which progressed to acute myeloid leukemia (AML) in some of the patients. Neurological and developmental findings were reported in one patient. We describe three patients in northern Israel from two unrelated families who reside in the same village. The first patient is a seven y.o girl, referred to genetic counseling due to mild developmental delay, behavioral disorders and pancytopenia. Bone marrow karyotype was normal. Exome sequencing revealed a homozygous splicing variant, c.3525+2T>G in *ERCC6L2*., predicted to damage splicing completely. The second patient is a 16 y.o boy who presented with MDS, progressed to AML after three months and passed away. Bone marrow demonstrated complex karyotype with 5qdel. Sequencing a panel of bone marrow disorders in fibroblasts derived from the patient revealed the same variant and an additional missense variant c.3493C>G p.Leu1156Val classified as vous. Segregation analysis exposed a 24 y.o healthy sister that was found homozygous for the variants. Carrier frequency testing within the village revealed 7 carriers out of 104 healthy individuals (1:15). There is a consensus among the hematological community that early hematopoietic stem cell transplantation (HSCT) is warranted for patients with MDS prior to the development of AML that is usually very aggressive.
Our data emerge the urgent identification of pre-symtomatic homozygous individuals to study the penetrance and the course of the disease in order to give accurate genetic and hematological counseling to the patients and family, and early treatment when indicated.

PrgmNr 2012 - The Immune-genetics landscape of brain metastases compared to their primary Tumors in a Saudi population

View session detail

Author Block: **D. Barakeh**[1,2], S. Abdulmohsen[2], L. Alhabeeb[3,2], S. S. Almohsen[4], D. Alayed[1], S. R. AlAnazi[5], M. Alzahrani[1], E. Alsolme[6], F. Alqubaishi[7], A. Almutairi[8], A. M. Binowayn[8], S. S. AlOtaibi[9], F. A. Alkhureeb[3], W. A. Shakweer[3], H. Al-Hindi[10], A. Alassiri[11], M. Abedalthagafi[8,9]; [1]Dept. of Pathology, King Saud Univ. Med. City, Riyadh, Saudi Arabia, [2]Genome Res. Ctr., King Fahad Med. City, Riyadh, Saudi Arabia, [3]Dept. of Pathology, King Fahad Med. City, Riyadh, Saudi Arabia, [4]Dept. of Pathology, King Abdulaziz Med. City, Riyadh, Saudi Arabia, [5]Dept. of Pathology, Security Forces Hosp., Riyadh, Saudi Arabia, [6]Genomics Res. Dept., King Fahad Med. City,, Riyadh, Saudi Arabia, [7]Genomics Res. Dept., King Fahad Med. City, Riyadh, Saudi Arabia, [8]Genomics Res. Dept., Saudi Human Genome Project, King Fahad Med. City, Riyadh, Saudi Arabia, [9]Saudi Human Genome Program, King Abdulaziz City for Sci. and Technology, Riyadh, Saudi Arabia, [10]Dept. of Pathology, King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia, [11]Dept. of Pathology and Lab. Med., King Abdulaziz Med. City, Riyadh,, Riyadh, Saudi Arabia

Disclosure Block: **D. Barakeh:** None.

Introduction: The most common central nervous tumors are metastatic tumors. Metastatic spread to the brain causes a significant drop in survival rates, and it usually occurs in patients with sufficiently controlled extracranial disease. Over half of the patients die within a few months following the diagnosis of brain metastasis. **Methods:** We performed Oncomine Panel Next Generation sequencing (161 cancer genes) of 68 matched brain metastases, and primary tumors from consented patients in multiple institutions in Saudi Arabia. We also performed *PD-L1* IHC to compare its expressions in paired primary tumors. **Results and discussion:** Rate of *PD-L1* positivity varies by metastatic location. Decisions for individualized therapies in patients with brain metastasis are often made from primary-tumor biopsies. We demonstrate that clinically actionable alterations present in brain metastases are frequently not detected in primary biopsies. Our results show patients with Breast Cancer who had *ERBB2* amplifications detected only in the brain-metastasis samples. Moreover, those patients have also had *EGFR V698M* mutations in the brain metastasis that are absent in their primary tumor. Activating mutations in *KRAS*, which have been associated with tumor responses to *MEK* inhibitors were the most frequent alteration in colon cancer paired. Further amplifications of *FGFR3* were detected only in the brain metastasis. The *MAPK pathway inhibitor* family includes agents that inhibit *BRAF* and *MEK*, such as vemurafenib, dabrafenib, or trametinib. *BRAF V600E* only detected in the brain-metastasis samples in our thyroid cancer cases. This suggests that sequencing of primary biopsies alone may miss a substantial number of opportunities for targeted therapy. Our preliminary results showed that the rate of *PD-L1* positivity varies by metastatic location. For example, *PD-L1* expression was rare on tumor cells in both the breast and metastatic sites. The difference in *PD-L1* positivity rates between primary tumors and different metastatic sites should inform physicians when choosing sites to biopsy and suggests a difference in the immune microenvironment across metastatic sites. **Conclusion:** Overall, the reason for these apparent organ site-specific differences is unclear. Because more than 50% of patients with brain metastases will die of intracranial progression, targetable alterations present in cancer subclones specific to the brain metastasis represent an important opportunity for novel targeted therapeutic strategies to affect overall survival.

PrgmNr 2013 - Timer CAR-T (TCAR-T): A paradigm shift in CAR-T cell therapy

View session detail

Author Block: **R. Roy**[1], P. KAUR[2,1], C. Mufeeda[1], S. Marakkar[1], K. W. Johnson[3], B. S. Glicksberg[4], L. Subramanian[5], K. Yadav[6], K. Shameer[7], A. Cherian[8]; [1]Molecular Robotics, Cochin, Kerala, India, [2]M. S. Ramaiah Univ. of Applied Sci., Bangalore, India, [3]New York, NY, [4]Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Courant Inst. of Mathematical Sci., New York Univ., New York, India, [6]Engineering Med., Texas A&M Univ., Houston, TX, [7]Mount Sinai Hlth.System, New York, NY, [8]Max Planck Inst. for Heart and Lungs Res., Bad Nauheim, Germany

Disclosure Block: **R. Roy:** None.

The development of effective targeted therapies with better survival and fewer side effects is crucial for cancer treatment. Although CAR-T cell therapy has shown immense potential in cancer treatment, off-target side effects, neurotoxicities, and cytokine release syndrome present unique challenges and have hindered its widespread use. Since CAR-T cells are designed to be alive in the body for their entire life span, it exacerbates the severe side effects. Thus, programming CAR-T cells to die after their assigned function can help overcome some of these challenges. Here, we propose the development of a new CAR-T approach to address these challenges. Briefly, we suggest that developing genetically modified CAR-T cells that can be eliminated using external modulations without compromising their therapeutic functions could address some of the shortcomings of the current cell therapy systems. These timer CAR-T (TCAR-T) cells can significantly act temporally, reduce long-term side effects, and be utilized during early tumorigenesis as they could be eliminated upon successful tumor remission. The development of such cells is possible by overexpressing the CAR protein and diphtheria toxin receptor together in the T-cells using a dual promoter virus vector. Since the modified TCAR-T cells contain comparatively more diphtheria toxin receptors, they could be highly sensitive and respond selectively to small doses of diphtheria toxin. The conditional removal of CART-T cells could also significantly reduce cytokine release during CAR-T cell therapy. We envisage that the validation TCAR-T approach could help ultra-personalize the cell and gene therapy offerings in the near future.

PrgmNr 2014 - A prospective Study Using Multigene Next Generation Sequencing panel to Evaluate the Prevalence of Familial Cancer in a Highly Consanguineous Population

View session detail

Author Block: **F. Alharthi**[1], M. Alharbi[2], N. Mobark[2], M. Abedalthagafi[1]; [1]King Abdulaziz City for Sci. and Technology, Riyadh, Saudi Arabia, [2]King Fahad Med. City, Riyadh, Saudi Arabia
Disclosure Block: **F. Alharthi:** None.
Introduction: 5-10% of all cancers are associated with an inherited mutation in a cancer predisposition gene. Saudi Arabia (SA), a highly consanguineous population has a sharp increase in cancer diagnosis to ~140% from 1990-2016. Family history is an important factor in determining cancer risk and prognosis. Understanding the population specific variations and frequencies is a major public health issue to establish informed screening programs. Here we conducted a prospective study using a multigene NGS panel and we established the first cancer genetic clinic in the SA.
Method: A prospective study was conducted between 2017-2020 on a total of 846 individuals. All ethical approval obtained. Cancer genetic clinic established and patients were seen by certified genetic counselors. A saliva or blood sample was obtained from 310 high risk patients including 166 index and 144 family members, and 536 controls. Cohort included 266 males and 580 females. gDNA sequenced using a 30 multigene panel important in hereditary cancers.
Results: Of the 310 subjects, 123 are index and 20 family members had previously been diagnosed with cancer, like breast, colon, brain. Out of these 310 subjects, 120 (38.7%) had positive genetic screening results for pathogenic or likely pathogenic mutations in APC, ATM, BRCA1, BRCA2, BRIP1, CDH1, CDKN2A, CHEK2, MLH1, MSH2, MSH6, MUTYH, NBN/NBS1, PALB2, PMS2, PTEN, RAD51D, and TP53. Variants of uncertain significance (VUS) were found in these genes, as well as in BARD1, BMPR1A, RAD51C, and STK11. No variants were found in BAP1, CDK4, EPCAM, GREM1, MITF, POLD1, POLE, or SMAD4. 24 of the controls also had pathogenic or likely pathogenic mutations in APC, ATM, BARD1, BRCA1, BRCA2, BRIP1, CDKN2A, EPCAM, MSH2, MSH6, PALB2, PMS2, including 3 individuals with more than one such mutation, and 260 had VUS, including 7 with pathogenic/likely pathogenic variants. Overall, there were 122 different variants in the 310 subjects, including 40 pathogenic or likely pathogenic variants, with many subjects having more than one variant, including having both pathogenic and likely pathogenic variants. In the control group, there were 378 different variants, but only 5 of these were also found in subjects.
Conclusion: Our study showed the high prevalence (38.7% in high risk, 5% in control, and 17% in average) of familial cancers in SA. Our data also support the importance of larger panel that contain a variety of hereditary cancer genes, even when patients lack classic clinical features or family history associated with some of the genes. Together, these findings further inform the clinical approach to hereditary genetic testing across a range of cancer types in SA populations.

PrgmNr 2015 - Genetic factors of differentiated thyroid cancer in French Polynesian population after exposure to nuclear tests, A suggested role of 3 loci

View session detail

Author Block: **M. Zidane**[1], M. Haber[2], F. RachÃ©di[3], J. Boissin[4], C. Ory[5], S. Chevillard[5], H. BlanchÃ©[6], R. Olaso[7], A. Boland[7], E. Conte[8], M. Karimi[1], T. Truong[1], Y. Ren[1], C. Xhaard[9], V. Souchard[1], J. Gardon[10], M. Taquet[10], J. F. DELEUZE[11], F. de Vathaire[1], J. Cazier[12]; [1]Inserm, Villejuif, France, [2]Wellcome Sanger Inst., Birmingham, United Kingdom, [3]Endocrinology Unit, Territorial Hosp. Taaone, Papeete, France, [4]Polynesian Inst. of Res. on Metabolic and Endocrine Disorders (IPRAME), Papeete, French Polynesia, Villejuif, France, [5]CEA, Fontenay aux roses, France, [6]CEPH, Paris, France, [7]CNRGH, CEA, Evry, France, [8]UPF, French Polynseia, France, [9]CHU Nancy, Nancy, France, [10]IRD, Montpellier, France, [11]CEA, EVRY, EVRY, France, [12]Univ. of Birmingham, Birmingham, United Kingdom

Disclosure Block: **M. Zidane:** None.

Background: Ionizing radiation (IR) exposure is the first established risk factor for differentiated thyroid cancer (DTC). However, most studies about DTC genetic factors in irradiated populations were performed in the Caucasian population, which was irradiated during the Chernobyl disaster, and did not evidence specific variants of genetic susceptibility to radiation-related DTC. Objective: To analyze the genetic factors of DTC risk among the French Polynesian population exposed to radioactive fallouts during the French nuclear tests. Methods: We studied 300,908 single nuclear polymorphisms (SNPs) genotyped in 283 DTC cases and 418 matched controls, all born in and resident French Polynesia, most of them being younger than 15 years at the time of the first nuclear test. We analyzed the genetic structure in our dataset to identify ethnic mixture subgroups. Then we completed a genome-wide analysis (GWAS) in the whole population, and also according to clinical features of the carcinoma. We performed local imputations for three genomic regions of interest showing significative association. We also performed gene-level and gene-set analyses subsequently.Results: We identified a strong genetic structure in the French Polynesian population related to admixture with Asian and European populations. GWAS results yielded three regions 10p12.2, 6q24.3, and 17q21.32 associated with DTC risk with respective p-values of the best three SNPs in these regions 8.40 x 10-8, 1.86 x 10-7, and 3.95 x 10-7. The respective allelic odds ratios (OR) were 1.95, 2.01, and 2.43. These results were similar in each ethnic admixture subgroup, but differentiated according to the clinical features of the thyroid carcinoma.To our knowledge, none of them have been associated previously with an increased risk of DTC. The gene-level analysis highlighted a probable interaction between RBFOX3 (17q25.3) and thyroid radiation dose in the risk of DTC.Conclusion: Our study suggests a novel role for the RBFOX3 and SKAP1 genes in addition to two genetic regions 10p12.2 and 6q24.3 in DTC risk.These results need to be further explored biologically.

PrgmNr 2016 - Hereditary Bilateral Breast Cancer: role of the Multi-Gene Panel Testing for detection of Pathogenic Variants

View session detail

Author Block: **C. Filorizzo**[1], D. Fanale[1], L. Incorvaia[2], A. Dimino[1], C. Brando[1], N. Barraco[1], M. Bono[1], K. Calcara[1], D. Cancelliere[1], L. Corsini[1], A. Cucinella[1], S. Cutaia[1], A. Fiorino[1], M. Greco[1], V. Gristina[1], J. Iovanna[3], G. Madonia[1], L. Magrin[1], E. Pedone[1], A. Perez[1], A. Pivetti[1], M. Ricciardi[1], R. Scalia[1], R. Sciacchitano[1], V. CalÃ²[1], S. Sammataro[1], G. Vaccaro[1], G. Badalamenti[1], V. Bazan[2], A. Russo[1]; [1]Dept. of Surgical, Oncological and Oral Sci., Section of Med. Oncology, Univ. of Palermo, Palermo, Italy, [2]Dept. of Biomedicine, NeuroSci. and Advanced Diagnostics (Bi.N.D.), Section of Med. Oncology, Univ. of Palermo, Palermo, Italy, [3]Ctr. de Recherche en CancÃ©rologie de Marseille (CRCM), INSERM U1068, CNRS UMR 7258, Aix-Marseille UniversitÃ© and Inst. Paoli-Calmettes, Parc Scientifique et Technologique de Luminy, Marseille, France
Disclosure Block: **C. Filorizzo:** None.
Bilateral breast cancer (BBC) is generally uncommon (1-2.6% of all patients with breast cancer), but its incidence increases particularly by up to 3% in *BRCA1* or *BRCA2* pathogenic variant (PV) carriers. The aim of our study was to evaluate whether all BBC patients should be offered multi-gene panel testing, regardless the criteria concerning personal and family history of cancer and age at diagnosis established by the current guidelines. We retrospectively collected and analyzed all clinical information of 150 BBC patients enrolled from October 2015 to April 2021, at the Sicilian Regional Center for the Prevention Diagnosis and Treatment of Rare and Heredo-Familial Tumorsâ of the Section of Medical Oncology of University Hospital Policlinico â P. Giacconeâ of Palermo. Recruited patients have been genetically tested for germline PVs in other gene beyond *BRCA1* and *BRCA2* by NGS-based multi-gene panel testing. In our investigation 58 (38.6%) out of 150 BBC patients harbored germline PVs in high and intermediate-penetrance breast cancer (BC) susceptibility genes, including *BRCA1*, *BRCA2*, *PTEN*, *PALB2*, *CHEK2*, *ATM*, *RAD51C*. Twenty-two out of 58 positively tested patients harbored a PV in a known BC susceptibility gene (no-*BRCA*). Interestingly, a noteworthy correlation between PVs in *PALB2* or *CHEK2* and BBCs was observed. In addition, our study showed that *CHEK2* PVs are correlated with a luminal A/B phenotype and *ATM* PVs with a luminal B subtype. Most of BBC-related *BRCA1/2* PVs were frameshift variants primarily located inside the exon 11 for *BRCA1*, and near the PALB2 binding site (at the N-terminus) and the DNA binding helical domain (at the C-terminus) for *BRCA2*. In conclusion, we found that, in the absence of an analysis performed via multi-gene panel, a significant proportion (14.7%) of PVs in genes different from *BRCA1/2* would have been lost. Our investigation led us to hypothesize that a deeper genetic analysis, through NGS-based multi-gene panel testing, could increase the detection rates of germline alterations in BBC patients. Particularly, in the near future, the evaluation of PVs could help to identify family members with a greater risk of developing BC (or other tumors) and consequently implement prevention and surveillance programs for these subjects. However, larger study cohorts are needed in order to define more accurate rates of PVs in BBC patients with previously negative *BRCA* genetic testing.

PrgmNr 2017 - Statistical Efficiency of Meta-analyses in Testing Gene-environment Interactions with Rare Variants

View session detail

Author Block: **X. Jin**, G. Shi; Xi'dian Univ., Xi'an, China
Disclosure Block: **X. Jin:** None.
Meta-analysis of results from multiple studies is a routine practice in genome-wide association studies, which offers a large effective sample size necessary for detecting associations with small effects. It is well established that the meta-analysis with common variants is statistically efficient and the power loss is minimal. However, the power loss of meta-analysis with rare variants is largely unexamined and has been suspected to be more sizable. In this work, we first propose four meta-analysis methods for testing gene-environment interaction (GÃE) effects with rare variants: HOM-INT-FIX, HOM-INT-RAN, HET-INT-FIX and HET-INT-RAN. Our methods consider homogeneous or heterogeneous GÃE effects across studies and treat the genetic main effect as either fixed or random. Then, we compare the statistical power of our meta-analyses with those of the pooled analyses that conduct a single interaction test with individual-level data from all studies. Through simulations, we show that HOM-INT-FIX and HOM-INT-RAN provide power comparable to those of the pooled analyses when the interaction effect is homogeneous across studies, which demonstrates that there is no power loss for the meta-analyses of GÃE with rare variants. On the other hand, when the interaction effect is heterogeneous across studies, HET-INT-FIX and HET-INT-RAN provide higher power than those of the pooled analyses. This is because that HET-INT-FIX and HET-INT-RAN treat genetic heterogeneity appropriately when synthesizing the summary results across studies. In contrast, the pooled individual-level data contain mixed interaction effects for the same variants, which violates the underlying assumption of the pooled analysis. Therefore, the meta-analysis of GÃE with rare variants is statistically efficient.

PrgmNr 2018 - DNA repair genes in polyposis susceptibility

View session detail

Author Block: **A. Olkinuora**[1], A. C. Mayordomo[2], A. K. Kauppinen[1], M. Cerliani[3], M. Coraglio[4], Ã. K. Collia[4], A. GutiÃ©rrez[4], K. Alvarez[5], F. LopÃ©z-KÃ¶stner[5], F. Jauk[6], H. GarcÃa-Rivello[6], A. RistimÃ¤ki[1], L. Koskenvuo[7], C. Vaccaro[2], W. H. Pavicic[2], P. T. PeltomÃ¤ki[1]; [1]Univ. of Helsinki, Helsinki, Finland, [2]Hosp. Italiano de Buenos Aires, Buenos Aires, Argentina, [3]Inst. Multidisciplinario de BiologÃa Celular, Buenos Aires, Argentina, [4]Hosp. Udaondo, Buenos Aires, Argentina, [5]Centro OncolÃ³gico, ClÃnica Univ. de Los Andes, Santiago, Chile, [6]Servicio de AnatomÃa PatolÃ³gica, Buenos Aires, Argentina, [7]Helsinki Univ. Hosp., Helsinki, Finland

Disclosure Block: **A. Olkinuora:** None.

Background: Adenomatous polyposis is a relatively common predisposing factor to colorectal cancer. Germline mutations in the *APC* gene underlie familial adenomatous polyposis (FAP). Up to 10% of patients with profuse disease (>100 polyps) and 80% of those with attenuated disease (AFAP) remain molecularly unexplained. While defective DNA repair may contribute to AFAP predisposition (e.g., germline mutations in the base excision repair genes *MUTYH* and *NTHL1*), repair defects are more common causes for other cancer syndromes, such as Lynch syndrome (DNA mismatch repair) or hereditary breast and ovarian cancer syndrome (DNA double strand break repair). We previously identified a homozygous founder variant of *MLH3* (c.3563C>G, p.Ser1188Ter) in five families with profuse or attenuated polyposis (PMID 30573798). Additionally, we found *MSH2* variants in two AFAP cases. To address the contribution of defective DNA repair to polyposis cases with the established susceptibility genes excluded, we scrutinized Finnish and South American cohorts with exome-wide and targeted methods.

Methods: Whole exome sequencing was conducted on 77 index cases and families with unexplained polyposis from the Helsinki University Hospital as well as nationwide Finnish and South American cancer registries. Variants with low population frequencies and predicted potentially pathogenic were selected. Co-segregation and mutational signature analyses were conducted whenever possible.

Results: We observed possibly pathogenic mono-and biallelic germline mutations in 59.7% of the patients. Several DNA repair genes were affected, most notably: *FANCM*, *MLH3*, *MSH2*, *MSH3*, *MUTYH*, *NTHL1*, *POLE* and *POLD1*. An additional polyposis family carrying the MLH3 (c.3563C>G, p.Ser1188Ter) founder variant was discovered. Strikingly, mutation signature analysis on the *MLH3* mutation carriers revealed a conserved signature 3 (defective homologous recombination-based DNA damage repair).

Conclusions: Overall, our results suggest that germline alterations in established and novel predisposition genes contributing to DNA repair may be present in a significant proportion of molecularly unexplained familial adenomatous polyposis cases.

PrgmNr 2019 - Bioinformatics analysis of gene expression profile in haematologic malignancies

View session detail

Author Block: **T. N. Patel**, P. Bhattacharya; Vellore Inst. of Technology, Vellore, India
Disclosure Block: **T.N. Patel:** None.
The study of gene expression profile in cancer enables the researchers to improve classification and characterization of tumors and define novel prognostic parameters. With development of microarray technology, the gene expression signatures have been thoroughly examined to translate this technology into â□□bench to bedsideâ□□. In the present study, we aimed to computationally analyze the differentially expressed genes (DEGs) in haematologic malignancies. The GEO (Gene Expression Omnibus) database was used to curate publicly available Affymetrix microarray datasets correspond to CML, AML, B-CLL, T-ALL, Hodgkinâ□□s and non-Hodgkinâ□□s lymphoma. The datasets were screened via GEO2R to identify the highly upregulated and downregulated genes for which |log (fold change)|â□¥1 and pFLT3, *MYC*, *HDAC9*, *RAD52*) were subjected to molecular docking with ATRA (All-trans retinoic acid) and ATRA like compound, Pumiliotoxin (Natural compound of *C. caesia Roxb.*). The overall results revealed three MMR genes- *MLH3* (upregulated), *MSH4*, and *PMS1* (downregulated) to be differentially expressed in AML, Hodgkinâ□□s lymphoma, and B-CLL respectively. The network analysis aided in identification of 12 DEGs associated with MMR through various pathways further depicting their functional interdependency in haematologic cancers. The enrichment analysis showed involvement of these genes in transcriptional regulation (positive and negative), DNA and transcription factor binding, and DNA repair. The molecular docking predicted that the overexpression of myc/flt3 and hdac9/rad52 can possibly be targeted by Pumiliotoxin and ATRA respectively. We conclude that the molecular perturbation of these 12 DEGs networked with MMR can pose detrimental impact on the genomic integrity and subsequently affect the functionality of mismatch repair. This will eventually accelerate the process of tumorigenesis in blood related cancers. Our study provides a novel insight on the role of â□□robust biomarker signatureâ□□ in modulating MMR in haematologic malignancies.

PrgmNr 2020 - Cancer informatics survey of different grades and subtypes of glioma

View session detail

Author Block: **P. KAUR**[1,2], R. Roy[2], K. W. Johnson[3], B. S. Glicksberg[4,5], L. Subramanian[6], K. Shameer[3], K. Yadav[7]; [1]M. S. Ramaiah Univ. of Applied Sci., Bangalore, India, [2]Molecular Robotics, Cochin, Kerala, India, [3]Mount Sinai Hlth.System, New York, NY, [4]Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Hasso Plattner Inst. for Digital Hlth.at Mount Sinai, Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Courant Inst. of Mathematical Sci., New York Univ., New York, NY, [7]Engineering Med., Texas A&M Univ., Houston, TX

Disclosure Block: **P. Kaur:** None.

Gliomas are brain tumors that originate in the glial cells. 'Glioma' is a blanket term used for defining different kinds of glial tumors, including astrocytoma, oligodendroglioma, and glioblastoma. The aggressiveness or malignancy of gliomas varies. While some are slow-growing and potentially treatable, others are invasive, proliferative, and tend to recur. This work aims to investigate the molecular trajectories of different grades of glioma to see how similar they are using a cancer informatics framework. We hypothesize that such a data-driven approach may aid in the discovery of new drug targets and miRNA that can potentially regulate tumor formation, which may benefit treatment and prognosis studies. We have used a translational bioinformatics approach to proceed with the investigation. We used the Disgenet database to retrieve canonical disease gene sets (18 diseases), including three WHO grade-wise tumors and eight WHO grade-wise subtypes of Diffuse astrocytic and oligodendroglia tumors, five other astrocytic tumors, and two high and low-grade tumor datasets. Enrichr-based pathway enrichment analysis revealed Viral myocarditis, Hepatitis B and C, Malaria, Measles, Influenza, Tuberculosis, and Alcoholism were enriched across different sub-types. STRING-db was used to compute protein networks using graph-based metrics, and critical genes such as *LRRC59, CCDC67, TRMT11,* and *TMEM135* were identified in Glioblastoma multiforme, while Cytoscape was used to prioritize candidates. Further evaluation of these targets could lead to identifying shared and distinct pathways associated with gliomas that could be targeted to develop new treatment approaches.

PrgmNr 2021 - Chromatin domains affect the clustered breakpoint formation in complex chromosomal rearrangement

View session detail

Author Block: **H. Inagaki**[1], T. Kato[1], A. Toyoda[2], Y. Makita[3], H. Kurahashi[4]; [1]Fujita Hlth.Univ., Toyoake, Japan, [2]Natl. Inst. of Genetics, MIshima, Japan, [3]Asahikawa Med. Univ. Hosp., Asahikawa, Japan, [4]ICMS, Fujita Hlth.Univ., Toyoake, Aichi, Japan
Disclosure Block: **H. Inagaki:** None.
The next generation sequencing technologies provide us novel views of chromosomes in the cells, such as interactions of the chromosomes, or chromosome conformations in nuclei. The rise of the long-read sequencing technology may offer further information of complex structural variations. We analyzed an autism patient carrying a complex chromosomal rearrangement using a long read sequencer. FISH analysis indicated rearrangements associated with chromosome 1, 2 and 14 as well as Yq12 repeated region. By using 30x high quality reads of around 20 kb library, Pacbio Sequel system successfully identified more than 40 breakpoint junctions. Yq12 shattered fragments as well as chromosome 2 and 14p fragments from ribosomal DNA repeats, ranging within 100 kb, were incorporated randomly into chromosome 1 and 2 breakpoints, representing a chromothripsis-like phenomenon. When the breakpoints were mapped onto the reference sequence, most breakpoints were located in gene bodies in both chromosomes 1 and 2. In addition, the breakpoints in chromosome 2 were found to be clustered in three regions, 2p22.1, 2p21, and 2p16.3. In each cluster, breakpoints were located within a single topologically associated domain, suggesting the fragmented breakage occurred in chromatin domain-dependent manner. On the other hand, such clustered manner was not found in chromosome 1 breakpoints. When the breakpoints in both chromosomes 1 and 2 were mapped with the ENCODE regulation tracks in UCSC web site, most of the breakpoints or breakpoints clusters were near the histone mark, H3K27Ac peaks, determined by ChIP-seq assays. Since the Yq12 repetitive region is specifically transcribed in male germ cells, it is suggested that the all fragmentation including Yq12 likely are transcription-dependent phenomenon. The results suggest that a subset of complex chromosomal rearrangements is related to the transcription, or chromatin states, possibly in spermatogenesis.

PrgmNr 2022 - Deep generative neural network for accurate drug response imputation

View session detail

Author Block: **P. Jia**[1], R. Hu[2], G. Pei[2], Y. Dai[2], Y. Wang[2], Z. Zhao[2,3]; [1]Key Lab. of Genomic and Precision Med., Beijing Inst. of Genomics, and China Natl. Ctr. for Bioinformation, Chinese Academy of Sci., Beijing, China, [2]Ctr. for Precision Hlth., Sch. of BioMed. Informatics, The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [3]Human Genetics Ctr., Sch. of Publ. Hlth., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Disclosure Block: **P. Jia:** None.

Drug response in cancer patients varies dramatically due to inter- and intra-tumor heterogeneity. While many studies have been conducted to identify signature genes or biomarkers to infer drug sensitivity or resistance, few have investigated the significant roles of transcriptome contexts in shaping the eventual treatment outcome. In this study, we developed a deep variational autoencoder (VAE) model to compress thousands of genes into latent vectors in a low-dimensional space which were subsequently used to build prediction models for a total of 251 compounds from two pharmacogenomics projects, Cancer Cell Line Encyclopedia (CCLE) and Genomics of Drug Sensitivity in Cancer (GDSC). The VAE models allowed non-linear compression of >6000 genes and thus, had the advantage to inform rich information across the whole transcriptome, such as the activities of various pathways, transcriptome contexts, and the relative expression levels of genes which might individually play minor roles. We applied rigorous quality assessment and validation to ensure the accuracy of the models in both cell lines and cancer samples. These included assessing the impact of cell line lineage, cross-validation, cross-panel evaluation, and validation in six independent clinical data sets. We demonstrate that using the embedding vectors could accurately impute drug response, outperform standard signature-gene based or linear-compression based approaches, and appropriately control the overfitting problem. Specifically, the component of the observed drug response that can be explained by gene expression achieved a high correlation across the CCLE and the GDSC panels for the 14 shared drugs. When applied to The Cancer Genome Atlas (TCGA) data, we showed that drugs with well-annotated targets confirmed the high quality of the predicted response, such as the ERBB2 inhibitor (Lapatinib), MET inhibitors (Crizotinib/PF2341066, Foretinib, and PHA-665752), BRAF inhibitors (PLX4720 and RAF265) and several MEK inhibitors. Furthermore, drug similarity was replicated using the predicted response both in CCLE cell lines and TCGA cancer samples. Finally, we demonstrated that drug response was associated with tumor microenvironment and tumor heterogeneity. In summary, our deep learning method and the results provided a useful resource to further explore genomic signatures that are associated with drug response.

PrgmNr 2023 - Distinctive miRNA expression profiling of pediatric solid tumors in relation to doxorubicin treatment

View session detail

Author Block: **C. Mujde**[1], S. Tug Bozdogan[1,2], E. Citak[3], B. Dogan[3], D. Karpuz[4], T. Erdem[5], A. Bisgin[1,2]; [1]Cukurova Univ., AGENTEM (Adana Genetic Diseases Diagnosis and Treatment Ctr.), Adana, Turkey, [2]Cukurova Univ., Faculty of Med., Dept. of Med. Genetics, Adana, Turkey, [3]Mersin Univ., Faculty of Med., Dept. of Pediatric Oncology, Mersin, Turkey, [4]Mersin Univ., Faculty of Med., Dept. of Pediatric Cardiology, Mersin, Turkey, [5]Mersin Univ., Faculty of Med., Dept. of Pediatrics, Mersin, Turkey
Disclosure Block: **C. Mujde:** None.
Introduction: microRNAs (miRNAs) are important posttranscriptional regulators consisting of 20-25 nucleotides that regulate hundreds of mRNA target genesâ expression. Each miRNA can act as oncogene or tumor supressor gene during the proliferation or oncogenesis. miRNAs are predicted to be potential prognostic/diagnostic biomarkers for various diseases in the last decade. Thus in this study we aimed to identify the expression profile of selected miRNAs in childhood tumors for its being the second cause of death between the ages of 5-14 and cardiotoxicity of treatment strategies such as doxorubicin. **Materials methods:** Thirty patients treated with doxorubicin in 2 groups of high and low dose (n=15 in each) who had childhood solid tumors and completed the treatment 2 years ago. Fifteen healthy age and sex matched children as a control group were also included. Forty miRNAs that had been selected due to metaanalysis in relation to disease perpetuation were studied. Total RNA were isolated, then the miRNA expression analysis were performed via Fluidigm, Biomark, USA. **Results:** Sixteen of 40 (40%) miRNAs (miR-126-3p, miR-142-3p, miR-142-5p, miR-148a-3p, miR-150-5p, miR-15a-3p, miR-182-5p, miR-18a-5p, miR-191-5p, miR-20b-5p, miR-223-3p, miR-23b-3p, miR-28-5p, miR-326, miR-423-5p, miR-93-5p) in patients receiving low-dose doxorubicin had a distinctive expression when compared to the control group. However, only 8 (20%) miRNAs (let-7b-5p, miR-150-5p, miR-191-5p, miR-20b-5p, miR-223-3p, miR-23b-3p, miR-342-3p, miR-93-5p) had a significant difference in high dose treated group. Additionally, 3 of 30 patients independent of doxorubicin dose had an ECG monitoring abnormalities. Most interestingly, only miR-191-5p expression was significantly decreased in these patients (p=0.031969). **Conclusion:** miRNA expression profiling is still a promising translational medical research area in relation to adverse events in different treatment modalities such as we present herein pediatric solid tumors.

PrgmNr 2024 - Exploration of hypomorphic variants from VUS in multigene panel testing for Japanese breast cancer patients

View session detail

Author Block: **C. Hata**[1], H. Nakaoka[2]; [1]Natl. Inst. of Genetics, Mishima, Japan, [2]Sasaki Inst., Chiyoda-ku, Japan

Disclosure Block: **C. Hata:** None.

The next generation sequencing technologies together with target enrichment systems dramatically reduce the cost to determine the sequences of multiple genes simultaneously. Thus, multigene panel testing allows us to detect pathogenic mutations that are not discovered by single gene testing. On the other hand, the use of multigene panel produces a large number of variants of uncertain significance (VUS) that have unknown functional effects and uncertain associations with cancer risk. We hypothesize that a small fraction of VUSs detected in multigene panel testing are hypomorphic variants that cause a partial loss of gene function and would be associated with cancer risk. In this study, we scrutinized hypomorphic variants in Japanese breast cancer patients. We used the target sequencing data for 11 breast cancer-related genes in 18,290 Japanese female individuals from Japanese Genotype-phenotype Archive (JGA, http://trace.ddbj.nig.ac.jp/jga), under accession number JGAS000140. The data contains 7,049 patients with breast cancer and 11,241 controls who were 60 years of age or older and did not have a past history nor family history of cancers. Single nucleotide variants and insertions and deletions were detected with the GATK. Variants were classified according to pathogenicity as follows: First, variants were classified as â□□benignâ□□ if their frequencies were greater than 1.0% in the public database in the 1000 Genomes or gnomAD project. Second, nonsense, splice site mutations, and indels were classified as â□□pathogenicâ□□. Third, we explored missense variants with previously established pathogenic or benign effects based on the ClinVar. Fourth, the other missense variants were classified as â□□VUSâ□□. Finally, we used frequency information in general populations and functional prediction scores by bioinformatics tools to detect candidate hypomorphic variants from the VUS.We confirmed that pathogenic variants in BRCA1/2 and PALB2 were significantly overrepresented in the patients with breast cancer. We selected hypomorphic variants from the VUS in 11 breast cancer-related genes. We showed that the detected hypomorphic variants in several genes were associated with breast cancer. The results suggest that a part of VUS have functional effects on cancer-related genes and are associated with the risk of developing breast cancer. Biological and biochemical assays to assess the impact of the candidate hypomorphic variants on protein function are needed.

PrgmNr 2025 - Germline landscape of BRCAs by 7-site collaborations as a BRCA consortium in Turkey

View session detail

Author Block: **A. Bisgin**[1,2], S. Ozemri Sag[3], M. Dogan[4], M. Yildirim[5], A. Aydin Gumus[6], N. Akkus[7], O. Balasar[8], C. Durmaz[9,10], R. Ersoz[11], S. Altiner[12,13], A. Alemdar[14], L. Aliyeva[3], I. Boga[1,2], F. Cam[6], B. Dogan[3,14], O. Esbah[15], A. Hanta[2], C. Mujde[2], C. Ornek[14], S. Ozer[2], C. Rencuzogullari[2], O. Sonmezler[2], S. Tug Bozdogan[1,2], M. Dundar[4], S. Temel[3,14]; [1]Cukurova Univ., Faculty of Med., Dept. of Med. Genetics, Adana, Turkey, [2]Cukurova Univ., AGENTEM (Adana Genetic Diseases Diagnosis and Treatment Ctr.), Adana, Turkey, [3]Bursa Uludag Univ., Faculty of Med., Dept. of Med. Genetics, Bursa, Turkey, [4]Erciyes Univ., Faculty of Med., Dept. of Med. Genetics, Kayseri, Turkey, [5]Necmettin Erbakan Univ., Meram Faculty of Med., Dept. of Med. Genetics, Konya, Turkey, [6]Manisa Celal Bayar Univ., Faculty of Med., Dept. of Med. Genetics, Manisa, Turkey, [7]Hlth.Sci. Univ., Kocaeli Derince Training and Res. Hosp., Dept. of Med. Genetics, Koaceli, Turkey, [8]Hlth.Sci. Univ., Konya City Hosp., Dept. of Med. Genetics, Konya, Turkey, [9]Hlth.Sci. Univ., Gazi Yasargil Training and Res. Hosp., Dept. of Med. Genetics, Diyarbakir, Turkey, [10]Hacettepe Univ., Faculty of Med., Dept. of Med. Genetics, Ankara, Turkey, [11]Duzce Univ., Faculty of Med., Dept. of Med. Genetics, Duzce, Turkey, [12]Hlth.Sci. Univ., Kanuni Training and Res. Hosp., Dept. of Med. Genetics, Trabzon, Turkey, [13]Ankara Univ., Faculty of Med., Dept. of Med. Genetics, Ankara, Turkey, [14]Bursa Uludag Univ., Inst. of Hlth.Sci., Dept. of Translational Med., Bursa, Turkey, [15]Duzce Univ., Faculty of Med., Dept. of Med. Oncology, Duzce, Turkey

Disclosure Block: **A. Bisgin:** None.

Introduction: *BRCA1/2* mutations play a significant role in cancer pathogenesis and predisposition particularly in breast, ovarian and prostate cancers. Thus, the detection of germline analysis of *BRCA1* and *BRCA2* is essential for clinical management strategies. Aiming at the identification of recurrent and novel mutations that could be used as a first screening approach. **Material Methods:** We analyzed next generation sequenced germline variants of *BRCA1/2* genes for 2168 individuals who had cancer diagnosis or high risk assessment due to BRCAs related cancers, referred to 10 health care centers distributed across 7 regions covering the all Turkish landscape. **Results:** Overall, 200 and 306 distinct mutations were identified in *BRCA1* and *BRCA2*, respectively. Twenty-five novel variants were reported from both genes while *BRCA2* showed higher mutational heterogeneity. We here in report the collective data as BRCA Turkish consortium that confirm the molecular heterogeneity in BRCAs among Turkish population while also as the first study sharing the both geographical, demographical and gene based landscape of all recurrent and novel mutations which some might be a founder effect in comparison to global databases. **Conclusion:** This wider perspective leads to the most accurate variant interpretations which pave the way for the more precise and efficient management affecting the clinical and molecular aspects.

PrgmNr 2026 - HIVID2: an accurate tool to detect virus integrations in the host genome

View session detail

Author Block: **X. ZENG**; Huazhong Agricultural Univ., Wuhan, China
Disclosure Block: **X. Zeng:** None.
Background Virus integration in the host genome is frequently reported to be closely associated with many human diseases, and the detection of virus integration is a critically challenging task. However, most existing tools show limited specificity and sensitivity. Therefore, the objective of this study is to develop a method for accurate detection of virus integration into host genomes. **Results and discussion** Herein, we report a novel method termed HIVID2 that is a significant upgrade of HIVID. HIVID2 performs a paired-end assembly (PE-assembly) for potentially integrated reads. The resulting sequences are then remapped onto the reference genomes, and both split and discordant chimeric reads are used to identify accurate integration breakpoints with high confidence. HIVID2 represents a great improvement in specificity and sensitivity, and predicts breakpoints closer to the real integrations, compared with existing methods. The advantage of our method was demonstrated using both simulated and real data sets. HIVID2 uncovered novel integration breakpoints in well-known cervical cancer-related genes, including *FHIT* and *LRP1B,* which was verified using protein expression data. In addition, HIVID2 allows the user to decide whether to automatically perform advanced analysis using the identified virus integrations. **Conclusions** By analyzing the simulated data and real data tests, we demonstrated that HIVID2 is not only more accurate than HIVID but also better than other existing programs with respect to both sensitivity and specificity. We believe that HIVID2 will help in enhancing future research associated with virus integration. HIVID2 can be accessed at https://github.com/zengxi-hada/HIVID2/.

PrgmNr 2027 - Mutation profile of circulating cell-free DNA reflects the evolution of metastatic breast cancer

View session detail

Author Block: **J. Kujala**[1], J. M. Hartikainen[1,2], M. TengstrÃ¶m[3], R. Sironen[1,3,4], P. Auvinen[3], V-M. Kosma[1,2,4], A. Mannermaa[1,2,5]; [1]Inst. of Clinical Med., Pathology and Forensic Med., Univ. of Eastern Finland, Kuopio, Finland, [2]Translational Cancer Res. Area, Univ. of Eastern Finland, Kuopio, Finland, [3]Cancer Ctr., Kuopio Univ. Hosp., Kuopio, Finland, [4]Dept. of Clinical Pathology, Kuopio Univ. Hosp., Kuopio, Finland, [5]Biobank of Eastern Finland, Kuopio Univ. Hosp., Kuopio, Finland
Disclosure Block: **J. Kujala:** None.
Background. About 20-30% of breast cancer (BC) patients with early-stage disease will eventually develop recurrent disease and have higher risk for developing metastatic BC. Risk for recurrence can be estimated with clinical and molecular tumor characteristics. However, predictive biomarkers are not uniformly presented in all cancer cells and intratumoral heterogeneity remains as a main cause of therapeutic failure. Patient blood samples contain small fractions of cell-free DNA (cfDNA) that originates from tumor cells and carries tumor-specific somatic variants. Liquid biopsy offers non-invasive and easily repeated method to assess the dynamic landscape of BC and may help to overcome challenges related to intratumoral heterogeneity. Here, we compared the performance of cfDNA-based liquid biopsy to matched tumor biopsies and provide further evidence for the use of liquid biopsy in the molecular profiling of early-stage BC. **Objectives-** To investigate how cfDNA reflects the heterogenic nature of early-stage BC and how differences in the mutational spectrum of cfDNA and tumor reflect the tumor evolution of recurrent BC. **Methods.** Two cohorts of patients with 1) locally advanced breast cancer and 2) early-stage BC who developed recurrent disease despite the good initial prognosis was selected for this study. cfDNA was isolated from patient serum samples and sequenced with targeted sequencing. Results were compared to matched tumors to estimate the concordance between samples. Tumor evolution was modeled with computational and deductive methods. **Results.** Somatic variants were detected especially in common BC associated genes such as *TP53*, *AKT1*, and *PIK3CA*. About 40% of detected variants were predicted as potential driver mutations and 7.5% of the variants as potential drug targets. Mean concordance between matched samples was 53.7%. Strong correlation between the tumor variant allele frequency (VAF) and cfDNA VAF was observed. Tumor evolution models illustrate the accumulation of somatic variants over time and share the characteristics of linear and branched tumor evolution. We were able to track recurrent disease back to primary tumor and detect recurrent-specific somatic variants in the cfDNA prior to secondary diagnosis. **Conclusions.** Somatic variants of BC tumors can be detected from the cfDNA at the time of diagnosis and prior to recurrence. The presence of somatic variants follows the progression of disease and allows the tracking of the tumor evolution when longitudinal samples are available. Discordance between the matched samples highlights the challenges related to liquid biopsy and supports the parallel sequencing of matched tumor and cfDNA.

PrgmNr 2028 - Reversal of Anthracycline Resistance in Colon Cancer cell lines using natural compound Curcumin

View session detail

Author Block: **S. Kanwal**[1,2]; [1]Univ. of the Punjab, Lahore., Lahore, Pakistan, [2]Sch. of Biochemistry and Biotechnology, Lahore, Pakistan
Disclosure Block: **S. Kanwal:** None.
The development of drug resistance is a major limitation of anthracycline usefulness in colon cancer.Multi drug resistance MDR reduces the effect of chemotherapeutics.MDR is mainly caused by overexpression of drug pumps that becomes a hurdle in chemotherapies. The present study aimed at the reversal of the epirubicin resistance in colon cancer cell line HCT116 EPI^{R} by curcumin as a reversal agent.Activity of the Curcumin was examined as a reversal agent by inhibiting the ABCC1 efflux pump by using HCT116 EPI^{R} 250nM cell line.The effective concentration of curcumin was found 5uM by MTT Assay and confirmed by Efflux assay with epirubicin via fluorescence microscopy.Drug accumulation in HCT116 EPI^{R} cells is confirmed when compared to their wild type counterparts and proved epirubicin is reversed by using curcumin. Further research is needed to understand the molecular basis of this resistance.

PrgmNr 2029 - Single cell transcriptomics of Pituitary Neuroendocrine Tumors (PitNETs)

View session detail

Author Block: **M. Brunner**, J. Meylan, A. Messina, M. Messerer, F. Santoni; CHUV, Lausanne, Switzerland
Disclosure Block: **M. Brunner:** None.
The pituitary gland, a master controller of hormones production and secretion driven by the hypothalamus, is a main component of the endocrine system. It is composed of five major cell types of different endocrine functions; somatotropes, corticotropes, lactotropes, gonadotropes and thyrotropes which produce growth hormone (GH), corticotropin (ACTH), prolactin (PRL), follicule stimulating hormone and luteinizing hormone (FSH, LH) and thyroid stimulating hormone (TSH) respectively. As in many other glands or organs, any individual or multiple cell types can abnormally differentiate into tumor cells. But unlike other organs, Pituitary Neuroendocrine Tumors (PitNETs) are almost exclusively adenomas (benign tumors). These tumors are classified based on 3 main characteristics; diameter (micro- or macro-adenoma Â± 1 cm), hormone secreting or non-secreting and invasiveness grade. PitNETs have been extensively characterized over the last decades with several methods; DNA sequencing to discover genome alteration differences between tumor types, Bulk RNA and DNA methylation in order to link those features with tumor invasiveness, prognosis and possible relapses. However, no single cell transcriptomic analysis of PitNETs have been published so far. Here, we present the results of a single cell analysis on our two first tumors, a plurihormonal (GH-PRL) tumor and a corticotropic (*POMC*) macro-adenoma. Characterization of the GH-PRL tumor showed heterogeneous cell populations; structural cells (endothelial and fibroblasts), immune cells (T and dendritic) and tumorigenic cells (GH+ and PRL+). Single cell RNA sequencing based copy number alteration analysis (CopyKat) revealed two distinct cell subset with different alteration pattern. Around 30% shows extensive alterations in chromosomes 9 and 13 while the rest exhibit few or no detectable alterations. Single cell characterization of the second tumor, mainly expressing *POMC* and *TBX19*, showed a major difference: the presence of a proliferating cluster of cells expressing *MKI67* and *TOP2A* markers. Intriguingly, *IQGAP3*, a gene already known to be implicated in different type of carcinomas, was highly expressed in this cluster, raising the question about common features between PitNETs and other types of tumors of endocrine origin. Pseudo-time analysis indicates a possible differentiation path from proliferating cells to *TBX19*+ cells passing by *POMC*+ cells. We planned more analysis of other PitNETs to help the interpretation of these findings and to give a new perspective on the comprehension of the structural composition and the dynamic progression of pituitary tumors.

PrgmNr 2030 - Spatial transcriptomics reveals unique molecular features of fluorescence sorted 5 minolevulinic acid positive infiltrative tumor cells associated with recurrence and poor survival in glioblastoma

View session detail

Author Block: **S. Chakraborty**[1], G. Andrieux[2], T. Das[1], S. Smith[3], R. Rahman[3]; [1]Univ. of Dhaka, Bangaldesh, Dhaka, Bangladesh, [2]Faculty of Med., Med. Ctr.-Univ. of Freiburg, Inst. of Med. Bioinformatics and Systems Med., Univ. of Freiburg, Freiburg, Germany., Freiburg, Germany, [3]Childrenâ□□s Brain Tumour Res. Ctr., Sch. of Med., Univ. of Nottingham, Nottingham, UK, Nottingham, United Kingdom
Disclosure Block: **S. Chakraborty:** None.
Spatiotemporal-heterogeneity of glioblastoma (GBM) originating from the genomic and transcriptional variation in spatially distinct sites, may contribute to subtype switching in GBM prior to and upon recurrence. Fluorescence-guided neurosurgical resection utilizing 5-aminolevulinic acid (5ALA) has enabled the isolation of infiltrative margin tumor cells (5ALA+ cells) from a background of non-neoplastic cells. We have explored the spatial-transcriptomic (ST) landscape to interrogate molecular signatures unique to infiltrating 5ALA+ cells in comparison to GBM core, rim, and invasive margin non-neoplastic cells. ST analysis reveals that GBM molecular subtype plasticity is not restricted to recurrence, but manifests regionally in a cell-type-specific manner. Whilst GBM core and rim are highly enriched with Classical and Proneural subtypes, the unique enrichment of the Mesenchymal subtype (MES) in 5ALA+ cells supports the hypothesis that MES 5ALA+ cells may drive GBM recurrence. Upregulation of the wound response pathway in 5ALA+ cells signifies the possibility of hijacking the wound healing pathway of neural cells to promote tumor growth. Exon-intron split analysis revealed an upregulation of exonic counts for MES and wound-response genes in 5ALA+ cells, implying that these genes are under active post-transcriptional control. Network analysis suggests that wound response genes, including chemokine *CCL2* that recruits regulatory T-cells and monocytic myeloid-derived suppressor cells, are controlled by an IRF8- mediated transcriptional program in 5ALA+ cells. A higher stemness signature both in 5ALA+ cells and non-neoplastic cells of the invasive margin emphasizes the role of this microenvironment in stemness acquisition and defines 5ALA+ cells as a rare sub-population of GBM stem cells. Finally, we establish a link between the unique molecular-signatures of 5ALA+ cells and poor survival and GBM recurrence. Characterization of the 5ALA+ infiltrative sub-population offers an opportunity to develop more effective GBM treatments and urges focus away from the GBM proliferative core, upon which failed targeted therapies have been predicated.

PrgmNr 2032 - *LPA*, KIV2, serum Lp(a), and the risk for Cancer and Cardiovascular disease related pathologies: A retrospective autopsy study

View session detail

Author Block: **M. Yamasaki**[1], M. N. Mieno[2], T. Arai[3], M. Sawabe[4], N. Tanaka[1]; [1]Dept. of Hlth.Data Sci. Res., Hlth.y Aging Innovation Ctr. (HAIC), Tokyo Metropolitan Geriatric Med. Ctr., Tokyo, Japan, [2]Dept. of Med. Informatics, Ctr. for Information, Jichi Med. Univ., Tochigi, Japan, [3]Dept. of Pathology, Tokyo Metropolitan Geriatric Hosp., Tokyo, Japan, [4]Dept. of Molecular Pathology, Graduate Sch. of Med. and Dental Sci., Tokyo, Japan

Disclosure Block: **M. Yamasaki:** None.

High Lp(a), lipoprotein (a), is known to be a risk factor of cardiovascular disease, while in cancer, low Lp(a) is reported to be a risk factor in our previous study. Lp(a) is composed of apolipoprotein (a) (apo(a)) and apolipoprotein B. Apo(a) contains highly polymorphic kringle domain, encoded by KIV2 (Kringle IV type 2) which is multiallelic copy number variation with about 5.5 kb sequence per copy and have 1 to > 40 copies per allele. High copy number of KIV2 shows low efficiency of membrane permeation and lead low Lp(a) concentration in blood. Heritability of Lp(a) is ~ 90% in European and African population and largely explained by *LPA* gene. Genetic study revealed association of KIV2 copy number and SNPs in *LPA* with cardiovascular disease. However, genetic association between cancer and KIV2 were not largely studied. Therefore, in this study we analyzed association of lifetime risk of cancer with Lp(a), SNPs in *LPA*, and KIV2 among autopsy samples in addition to lifetime risk related to cardiovascular disease. Participants were 304 autopsies from the Tokyo Metropolitan Geriatric Hospital. Lp(a) concentration was measured from fresh blood of the participants using a latex- enhanced turbidimetric immunoassay. DNA was extracted from the renal cortex by the phenol-chloroform method and stored frozen until use. Exome array were performed previously. To measure KIV2 copy number, droplet digital PCR (ddPCR) for these samples were conducted. Following pathological outcomes were used in our analysis; cancer, myocardial infarction, ischemic heart disease, brain infarction, cerebrovascular disease, degree of arteriosclerosis and intracranial atherosclerotic index. In the presentation, we will report association of Lp(a) with SNPs in *LPA* region and KIV2 copy number. In addition, analysis of above outcomes and with Lp(a), SNPs in *LPA* region and KIV2 copy number, adjusting sex, age at death, smoke, alcohol, diabetes, BMI, HDL-C and total cholesterol, will be performed.

PrgmNr 2033 - Trans-ancestry GWASs for ECG markers of ventricular depolarization and repolarization in 250,730 individuals identifies shared and distinct mechanisms

View session detail

Author Block: **W. Young**[1], A. Isaacs[2], N. Lahrouchi[3], B. Mifsud[4], P. B. Munroe[5], C. H. Newton-Cheh[6], N. Sotoodehnia[7], CHARGE consortium EKG working group; [1]Queen Mary Univ. of London, London, United Kingdom, [2]Univ. of Maastricht, Maastricht, Netherlands, [3]Univ. of Amsterdam, Amsterdam, Netherlands, [4]Hamad Bin Khalifa Univ., Doha, Qatar, [5]Queen Mary Univ. of London, London, United Kingdom, [6]Massachusetts Gen Hosp, Boston, MA, [7]Univ. of Washington, Seattle, WA
Disclosure Block: **W. Young:** None.
Background: The QT interval is an important ECG measure of cardiac ventricular electrophysiology. It is the sum of QRS duration and JT interval, which are markers of ventricular depolarization (VD) and repolarization (VR) respectively. Therefore, analyses of QRS and JT give insight into each phase of ventricular electrophysiology. These ECG traits are heritable and are independent predictors of ventricular arrhythmia and sudden cardiac death (SCD). Previous genome-wide association studies (GWAS) for QRS and JT highlighted pathways regulating cardiac ion channels, calcium signalling, myocardial mass and myocyte internal structure. Their heritability, however, remains largely unexplained. We performed the largest GWAS to date for QRS and JT to identify new candidate genes and elucidate novel mechanisms underlying arrhythmogenesis relating to VD and VR. Methods: This trans-ancestry GWAS, using 35 studies imputed with 1000 Genome / HRC reference panels, comprised a total sample of 252,730 individuals (84% European, 7.7% Hispanic and 6.7% African ancestry). Candidate gene prioritisation and gene-set enrichment analyses were performed using DEPICT. Results: We identified 155 and 121 independent loci for JT and QRS, of which 96 and 77 were novel. 41 loci overlapped and at 34 of these, the lead variants were genome-wide significant in both analyses. Discordant directions of effect were observed at 27 (79.4%) of these loci. Lead variants explained approximately 54% and 42% of the heritabilities of JT and QRS. A negative genetic correlation was observed between the traits (r_g = -0.25, P = 0.003). Gene-sets for JT and QRS were highly expressed in cardiac tissues. Additionally, for QRS, connective tissue cell-types, including cardiac valves and chondrocytes, were significant. Common top Gene-Ontology terms included biological processes involved in cardiac/muscle cell differentiation, muscle tissue development and embryonic development (FDR GLUT4, *PIGQ* and *ABCC8*. For QRS, heart/organ growth and transmembrane receptor protein serine/threonine kinase signalling pathway regulation were significant, including genes *GATA4* and *ID2,* which encode a cardiac transcription factor and a transcription regulator. Conclusions: Our analyses highlight novel genes and processes for VD and VR, with shared and distinct mechanisms. These findings may expose new pathways which contribute to arrhythmogenesis and SCD, and could serve as new therapeutic targets.

PrgmNr 2034 - A Novel Method of Identifying Inherited High-risk Population of Atherosclerotic Cardiovascular Disease Using Polygenic Risk Scores of Metabolic Diseases

View session detail

Author Block: **H. Song**[1], Y. Koh[2], T-M. Ree[2], S-Y. Choi[2], S. Kang[3], S-P. Lee[2]; [1]Genome Opinion, Inc., Seoul, Korea, Republic of, [2]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, [3]Gangnam Severance Hosp., Seoul, Korea, Republic of
Disclosure Block: **H. Song:** Major Stockholder/Ownership Interest; Genome Opinion, Inc.. Salary/Employment; Genome Opinion, Inc.. Receipt of Intellectual Property Rights/Patent Holder; The content of this work has been filed as a patent.
Atherosclerotic cardiovascular disease (ASCVD) is the leading cause of death worldwide and the development of methods to identify high risk population of ASCVD is an important health issue. According to recent studies, the polygenic risk score (PRS) could predict ASCVD risk by use variants associated with the disease itself. However, we do not yet know that how to use polygenic risk of metabolic diseases to predict inherited risk of ASCVD. In this study, we established metabolic disease PRSs and to demonstrate its clinical utility in predicting incident ASCVD using over 40,000 Korean genotypes. We constructed the PRSs for eight quantifiable metabolic phenotypes - systolic/diastolic blood pressure, body mass index, four blood lipid components, and fasting blood glucose - from genome-wide association studies of derivation cohorts (n=37,285) and conducted a grid search of combination of these metabolic disease PRSs to identify the optimal set and weighted score for incident ASCVD prediction. In the independent validation cohorts (n=4,333), the participants in the 5th quintile of each metabolic disease PRS show 1.4~1.9-fold increased risk of incident hypertension, obesity, hyperlipidemia, and diabetes. Using combined metabolic disease PRS model, we identified 6.7% of the population as a high risk group demonstrating a 3.3-fold (1.7-6.1, 95% C.I.; P

PrgmNr 2035 - A translational bioinformatics investigation of the human gut microbiome and hypertensive diseases

View session detail

Author Block: **A. Prasannakumar**[1,2], R. Roy[3], K. W. Johnson[4], B. S. Glicksberg[5], L. Subramanian[6], K. Yadav[7], N. Chia[8], I. J. Kullo[9], J. Dudley[4,10], K. Shameer[4]; [1]Union Christian Coll., Aluva, Kerala, India, [2]Molecular Robotics, Cochin, Kerala, India, [3]Molecular Robotics, Kerala, Kerala, India, [4]Mount Sinai Hlth.System, New York, NY, [5]Hasso Plattner Inst. for Digital Hlth.at Mount Sinai, Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Courant Inst. of Mathematical Sci., New York Univ., New York, NY, [7]Engineering Med., Texas A&M Univ., Houston, TX, [8]Microbiome Program, Ctr. for Individualized Med., Mayo Clinic, Rochester, MN, [9]Mayo Clinic, Rochester, MN, [10]Onegevity, New York, NY
Disclosure Block: **A. Prasannakumar:** None.
Understanding the complex relationship between the human gut microbiota and cardiovascular disease is emerging and of particular interest as diet plays a critical role in the severity and outcome of cardiovascular diseases. We hypothesize that understanding host-microbe dynamics could help understand novel disease pathways and help develop microbiome-driven therapeutics. This study explores the role of human gut microbiota in individuals with hypertension using a translational bioinformatics analytics framework involving meta-analyses, gene enrichment analysis, cross-genome metabolic pathway comparisons, and drug repositioning. Eight major hypertensive disorders, namely, diastolic hypertension, elderly hypertension, grade 3 hypertension, pulmonary artery hypertension, portal hypertension, systolic hypertension, preeclampsia, and primary hypertension, were selected. We linked important microbial genera to these disorders using a data discovery approach. We identified 12 taxa (*Bacteroides, Blautia, Faecalibacterium, Bifidobacterium, Escherichia, Prevotella, Streptococcus, Eubacterium, Akkermansia, Parabacteroides, Roseburia,* and *Clostridium*) using data compiled in MuPhenome. According to our findings, 915 human genes were associated with eight hypertensive diseases. We discovered 75 metabolic pathways shared by these microorganisms and distinct hypertensive disorders. Pathways co-modulated by host and microbes implicated in hypertensive genes include amino-acid metabolism, mannose degradation, glycerolipid metabolism, nucleotide metabolism, metabolism of tRNA, folate metabolism, and fatty acid beta-oxidation. Few drugs were found to be effective against the three genomic targets, namely *PAH* (Sapropterin), *PDE5A* (Sildenafil, Tadalafil, Udenafil, and Vardenafil), *PDGFRB* (Sorafenib), XDH (Allopurinol), which are suggesting potential drug repositioning opportunities. Further research focusing on co-metabolic pathways and genes implicated in host-microbe interaction might aid in developing gut-microbiome-based therapies against hypertensive diseases.

PrgmNr 2036 - Genome-wide association study of lipids in Greenlandic Inuit show large-effect size and independent associations and reveals a unique genetic architecture

View session detail

Author Block: **N. Senftleber**[1,2], E. JÃ¸rsboe[3], M. Andersen[3], G. Garcia-Erill[2], J. Meisner[2], N. Grarup[3], A. Gilly[4], P. Bjerregaard[5], C. Larsen[5,6], M. JÃ¸rgensen[7], E. Zeggini[4], I. Moltke[2], T. Hansen[3,8], A. Albrechtsen[2]; [1]Steno Diabetes Ctr. Copenhagen, Lyngby, Denmark, [2]Bioinformatics Ctr., Dept. of Biology, Univ. of Copenhagen, KÃ¸benhavn, Denmark, [3]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Faculty of Hlth.and Med. Sci., Univ. of Copenhagen, KÃ¸benhavn, Denmark, [4]Inst. of Translational Genomics, Helmholtz Zentrum MÃ¼nchen â German Res. Ctr. for Environmental Hlth., Neuherberg, Germany, [5]Natl. Inst. of Publ. Hlth., Univ. of Southern Denmark, KÃ¸benhavn, Denmark, [6]Greenland Ctr. for Hlth.Res., Univ. of Greenland, Nuuk, Greenland, [7]Steno Diabetes Ctr. Greenland, Nuuk, Greenland, [8]Faculty of Hlth.Sci., Univ. of Southern Denmark, Odense, Denmark

Disclosure Block: **N. Senftleber:** None.

Objective: Perturbation of the lipid homeostasis is a key risk factor for cardiovascular disease (CVD), the leading cause of death worldwide. We aimed to investigate the role genetics plays for lipid levels, and via these CVD, in Greenlanders. **Methods:** Using imputed data from 4473 Greenlanders, we performed a genome-wide association study (GWAS) of levels of six blood lipids and lipoproteins: triglycerides, LDL-cholesterol, HDL-cholesterol, total cholesterol, apolipoprotein A1, and apolipoprotein B. For relevant genome-wide significant variants, we then tested for associations with additional traits and outcomes, including risk of CVD and RNA expression. Finally, we investigated the combined effect of the independent genome-wide significant variants using genetic risk scores (GRSs). **Results:** The GWAS identified 11 genome-wide significant loci, of which 9 were known. Two were potentially novel, however, they were more frequent in Europeans and not Bonferroni significant and therefore need independent replication to be validated. In the *PCSK9* locus, we identified an association signal, which was independent of the well-described *PCSK9* loss-of-function variant rs11591147. The derived allele of the lead SNP, rs4927191, in this locus was associated with lower levels of LDL-cholesterol (beta(SD)=-0.226, p=1.24e-12) and total cholesterol (beta(SD)=-0.169, p=2.28e-9). Interestingly, it was also associated with a lower risk of peripheral artery disease (hazard ratio=0.30 per allele, p=0.01), and was a top eQTL signal for *PCSK9* across all tissues in GTEx. A variant (rs12117661) in LD with the lead SNP showed similar results, and there were furthermore indications of rs12117661 being located in a predicted regulatory element. When combining the GWAS hits into GRSs using 10-fold cross-validation, these scores alone explained up to 16.3% of the variance of the lipid traits. Furthermore, the GRSs for LDL-cholesterol, total cholesterol and apolipoprotein B were associated with increased risk of several types of CVD events. **Conclusion:** We identified a possible causal variant that affected lipid levels and CVD risk by regulating the expression of *PCSK9*. Furthermore, we found that in the Greenlandic population much more of the variance in lipid levels can be explained with markedly fewer variants than in large European populations, supporting a marked difference in genetic architecture.

PrgmNr 2037 - Integration of biomarker polygenic risk score improves the prediction of coronary artery disease in UK Biobank and FinnGen

View session detail

Author Block: **J. Lin**[1], N. Mars[1], Y. Fu[1], J. P. Ripatti[1], FinnGen, S. Ripatti[1,2,3], M. Pirinen[1,2,4]; [1]Inst. for Molecular Med. Finland, Helsinki, Finland, [2]Publ. Hlth., Helsinki, Finland, [3]Broad Inst. of Massachusetts Inst. of Technology and Harvard Univ., Cambridge, MA, [4]Dept. of Mathematics and Statistics, Helsinki, Finland

Disclosure Block: **J. Lin:** None.

Smoking history, age, levels of blood pressure, cholesterol, lipoproteins and inflammation are established biomarkers for coronary artery disease (CAD). As canonical polygenic risk scores (PRS) have recently proven successful for CAD prediction, it remains of high interest to determine to what extend a meta-biomarker score (metaPRS) constructed from statistically relevant biomarkers can contribute upon the established CAD PRS predictions. We derived metaPRS scores using SNPs selected from 16 genome-wide association studies of CAD related biomarkers from the UK Biobank (UKB) with 322,058 (5.45% cases) white, unrelated subjects, subsampled into training and validation sets. Statin-adjusted optimal biomarker subsets for incident CAD were identified via data regularization using glmnet. The SNPs of interest were selected using PRS-CS (Ge et al.) and LD-pruning and p-value clumping and then integrated with the canonical genome-wide PRS score from Khera et al. (KheraBiomPRS). These scores were tested on the UK Biobank data and on 306,287 (10.04% cases) subjects within the Finnish FinnGen study. Within the UKB training set, we constructed and contrasted metaPRS scores from a 10 biomarker (optimal fit; 25,677 SNPs) and 15 biomarker (minimum cross-validation (mincv) error; 33,476 SNPs), mincv PRS-CS (1,103,911 SNPs) models and tested them on UKB validation and FinnGen cohorts. We found that PRS-CS biomarker PRS models yielded the highest adjusted variance explained (R^2) scores. Within the training set, the hazard-ratio (HR) for KheraPRS was 1.64 (1.61-1.67) per one standard deviation, and for KheraBiomPRS HR was 1.68 (1.64-1.73) for the optimal model 1.69 (1.66-1.73) for the mincv model and 1.72 (1.69-1.75) in the PRS-CS model. Within the FinnGen study, the HR for KheraPRS was 1.53 (1.51-1.55), and the HR for KheraBiomPRS was 1.55 (1.53-1.57) for both the optimal and mincv models while the HR for PRS-CS model was 1.57 (1.55-1.59). Comparing the 99% quantile against the 40-60% reference group in UKB, KheraPRS had HR of 4.39 (3.93-4.91) while the mincv model had HR of 4.80 (4.30-5.35), compared with 4.79 (4.29-5.35) for the PRS-CS model. In FinnGen, the corresponding 99% quantile HR values were 3.42 (3.16-3.70) for KheraPRS while KheraBiomPRS had HR 3.45 (3.19-3.73) for the mincv model and increased to 3.63 (3.36-3.92) for the PRS-CS model. The largest KheraBiomPRS HR gain of 24% was found in the early CAD onset (age Integration of meta-biomarker PRS improves on the existing canonical CAD PRS predictions although the gain might be smaller across populations.

PrgmNr 2038 - Multivariate genetic analysis of human plasma glycerophospholipids, glycerolipids, sphingolipids and sterols identifies novel associations near *LPGAT1*, *GRIP1*, *SGPL1* and *ERMP1*

View session detail

Author Block: **L. Ottensmann**[1], R. Tabassum[1], S. E. Ruotsalainen[1], M. J. Gerl[2], E. WidÃ©n[1], K. Simons[2,3], FinnGen, S. Ripatti[1,4,5], M. Pirinen[1,5,6]; [1]Inst. for Molecular Med. Finland, HiLIFE, Univ. of Helsinki, Helsinki, Finland, [2]Lipotype GmbH, Dresden, Germany, [3]Max Planck Inst. of Molecular Cell Biology and Genetics, Dresden, Germany, [4]The Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Dept. of Publ. Hlth., Clinicum, Faculty of Med., Univ. of Helsinki, Helsinki, Finland, [6]Dept. of Mathematics and Statistics, Univ. of Helsinki, Helsinki, Finland
Disclosure Block: **L. Ottensmann:** None.
The human plasma lipidome, composed of thousands of lipid species, captures information beyond the routinely clinically used lipids and have disease relevance in cardiometabolic diseases and beyond. Genome-wide association studies (GWAS) of individual lipid species (univariate analysis) have identified many lipid-associated loci, however the influence of genetic variants on lipid metabolism and cardiovascular disease risk is still not fully understood. Multivariate analysis of multiple correlated lipid species improves statistical power and could help identify additional lipid loci. We performed univariate and multivariate analysis using lipidome data.

We conducted univariate GWAS of 11.3 M variants and 7177 participants from the Finnish GeneRisk cohort with the software MMM for 104 glycerophospholipids, 44 glycerolipids, 15 sphingolipids and 16 sterols and sterol esters. We grouped the lipid species into 11 clusters based on their correlations and performed multivariate analyses for each cluster by metaCCA software. The univariate and multivariate associations were fine-mapped with FINEMAP to identify causal SNPs underlying the associations. The putative causal variants were examined for associations with cardiometabolic traits in FinnGen (Release 6), Gene Atlas (UKBB) and GWAS Atlas.

Multivariate analysis across 11 clusters identified 79 loci with P-value 3 Mb apart from each other. In addition to well-known lipid loci such as *FADS1-2-3*, *CPT1A* and *NTAN1*, we identified four new loci *LPGAT1* (*rs143208479*, *rs28432809*), *GRIP1* (*rs118025056*), *SGPL1* (*rs12763964*) and *ERMP1* (missense variant *rs140094646*), at a Bonferroni-corrected significance threshold of P-value SGPL1 reached the Bonferroni-corrected significance threshold in a univariate analysis (Cer42:2;2) highlighting the additional statistical power provided by multivariate association test. Additionally, in the univariate analysis, a novel association near *AP005242.1* (*rs554083837*) was identified for four phosphatidylcholine species. SNPs *rs143208479* and *rs28432809* near *LPGAT1* are associated with ulcerative colitis with primary sclerosing cholangitis in FinnGen (P=2e-5) and mean reticulocyte volume in Gene Atlas (P=2e-9) respectively.

Multivariate genetic analysis is a powerful tool for high-dimensional data such as lipidomics and helped us identify four novel lipid-associated loci: *LPGAT1*, *GRIP1*, *SGPL1* and *ERMP1*, of which only the locus *SGPL1* was also identified by the univariate analysis.

PrgmNr 2039 - Rare Variant Analysis for Coronary Heart Disease Cases and Controls using Whole Genome Sequencing in a Middle Eastern Population Identifies Potential Novel Genes

View session detail

Author Block: **E. Ullah**[1], K. Kunji[1], A. El-Menyar[2], I. J. Kullo[3], M. Saad[1], J. Al Suwaidi[2]; [1]Qatar Computing Res. Inst., Hamad Bin Khalifa Univ., Doha, Qatar, [2]Hamad Med. Corp., Doha, Qatar, [3]Dept. of Cardiovascular Med., Mayo Clinic, Rochester, MN
Disclosure Block: **E. Ullah:** None.
Background: Rare variants can contribute to the missing heritability of complex diseases such as coronary heart disease (CHD). With advances of next generation sequencing and their decreasing cost, the analysis of rare variants has become more feasible. Rare variant analysis can provide new insights to the etiology of complex diseases.
Methods: In a Middle Eastern cohort from the Qatar Cardiovascular Biorepository, Qatar Genome Programme, and Qatar Biobank, whole genome sequencing (WGS) with 30x coverage was conducted on coronary heart disease patients (n=1,067, mean age Â± SD = 59.96 years Â± 10.99; 70.32% males) and controls (n=6,170, mean age Â± SD = 40.02 years Â± 12.56; 43.45% males). Burden and SKAT gene-based analyses were performed via a linear mixed model accounting for age, sex, BMI, and principal components using SMMAT. We included rare variants within genes with MAF **Results:** All statistical distributions were well controlled. Using a Bonferroni threshold (*P*-6), a total of 10 and 15 genes were significant using burden and SKAT analyses, respectively. These genes are *SLC22A5* (2 variants, burden $P=1.14x10^{-13}$, SKAT $P=3.79x10^{-18}$), *RP11-468E2.6* (2 variants, burden $P=1.22x10^{-9}$, SKAT $P=1.94x10^{-10}$), and *WBSCR22* (3 variants, SKAT $P=9.42x10^{-7}$). Mutations in *SLC22A5* can cause carnitine deficiency and the gene is a regulator of lipid metabolism. Common SNPs within this gene were also associated with low density lipoprotein cholesterol. *WBSCR22* has been associated with Williams-Beuren Syndrome, whose most significant medical problem is cardiovascular diseases caused by narrowed arteries. Among the less significant genes, we identified *APOB* (4 variants, burden $P=1.22x10^{-4}$, SKAT $P=6.04x10^{-5}$) and *NUP93* (2 variants, burden $P=1.78x10^{-3}$, SKAT $P=1.44x10^{-4}$). SNPs near *NUP93* have been associated with high density lipoprotein cholesterol, and ApoA1 and ApoB measurements.
Conclusions: Rare variant analysis has identified novel genes with potential biological link to CHD in a Middle Eastern WGS cohort. Further studies are needed to validate these genes.

PrgmNr 2040 - Sex-stratified whole genome sequencing association study for coronary heart disease in a middle eastern population suggests differential loci

View session detail

Author Block: **K. Kunji**[1], E. Ullah[1], A. El-Menyar[2], I. J. Kullo[3], M. Saad[1], J. Al Suwaidi[2]; [1]Qatar Computing Res. Inst., Hamad Bin Khalifa Univ., Doha, Qatar, [2]Hamad Med. Corp., Doha, Qatar, [3]Dept. of Cardiovascular Med., Mayo Clinic, Rochester, MN
Disclosure Block: **K. Kunji:** None.
Introduction: There has been a dearth of genome-wide association studies (GWASs) in Middle Eastern populations for complex diseases such as coronary heart disease (CHD). CHD diagnosis and progression are male/female differential. There are only a few sex-stratified CHD GWASs, which were performed in European ancestry cohorts.
Hypothesis: We hypothesize that there are differences in the genetic etiology of CHD between males and females.
Methods: We performed a sex-stratified GWAS on CHD using whole genome sequencing data in a Middle Eastern population consisting of 713 male CHD patients and 2,588 male controls (mean age Â± SD = 58.75 yrs Â± 11.54 and 39.76 yrs Â± 12.18, respectively) and 301 female CHD patients and 3,368 female controls (mean age Â± SD = 59.46 yrs Â± 9.54 and 40.22 yrs Â± 12.85, respectively). Association testing was conducted for the autosomal and X chromosomes in each sex separately using linear mixed models. Fisherâ□□s meta-analysis was performed on the X chromosome to identify SNPs associated in both sexes. We assessed the performance of a polygenic risk score, metaGRS, for each sex.
Results: Previously reported loci/genes were replicated in our analysis at P CDH13, *MIR548A3*, *PRKG1*/*A1CF*, and *REST* in females; *TRIM5*, *ARHGAP42*, *ILR6*, and *MYO9B* in males). We also found suggestive loci (P PCSK5 and *PTGS1* in females; *HRC*, *CSNK2A1* and *PLCB1* in males). Area under the receiver operating curve of metaGRS was similar between the sexes (~0.68). However, OR per 1 SD increase was lower in females (1.46 [1.29 - 1.64]) than in males (1.78 [1.58 - 1.99]). The top Fisherâ□□s SNP is by *P2RY4*, EDA (binds *EDA2R*, prior association).
Conclusions: Our study replicates different known CHD loci in females and males and suggests new genes. Good performance of metaGRS was observed for both sexes, but the lower OR per 1 SD in females suggests a stronger result for males.

PrgmNr 2041 - The penetrance of rare cardiomyopathy-associated DNA variants: a cross-sectional approach to estimate penetrance using large case series

View session detail

Author Block: **K. McGurk**[1], X. Zhang[2], P. Theotokis[1], K. L. Thomson[3], A. R. Harper[4], Cardiac Variant Interpretation Consortium, M. Farrall[4], H. Watkins[3], L. Bottolo[5], J. S. Ware[6]; [1]Natl. Heart and Lung Inst., Imperial Coll. London, London, United Kingdom, [2]European Bioinformatics Inst., Hinxton, United Kingdom, [3]Radcliffe Dept. of Med. and the Wellcome Trust Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom, [4]Radcliffe Dept. of Med. and the Wellcome Trust Ctr. for Human Genetics, Univ. of Oxford, London, United Kingdom, [5]Dept. of Med. Genetics and the MRC Biostatistics Unit, Univ. of Cambridge and the Alan Turing Inst., Cambridge, United Kingdom, [6]Natl. Heart and Lung Inst., Imperial Coll. London; MRC London Inst. of Med. Sci.; Cardiovascular Res. Ctr. at Royal Brompton and Harefield Hosp., London, United Kingdom
Disclosure Block: **K. McGurk:** None.
Inherited cardiomyopathies (CM) are characterised by clinical and genetic heterogeneity, incomplete and age-dependent penetrance, and variable expressivity. Understanding the penetrance of disease-causing variants identified in families with disease or as secondary findings in the population is of paramount importance. Genes associated with CM account for 23% of the ACMG73 genes, where opportunistic screening for secondary findings is recommended, yet our understanding of penetrance for an individual â□□pathogenicâ□□ variant found in this setting is extremely limited.Current longitudinal cohorts may be informative for rare variants aggregated within a gene, but data for individual variants remain sparse given rarity. To address this, we estimated variant-specific penetrance for rare CM-associated variants through aggregation of data from large multi-centre case cohorts comprising >10,000 hypertrophic CM (HCM) cases and 2,300 dilated CM (DCM) cases, using a cross-sectional approach that compares variant frequencies between affected individuals and reference populations (UK Biobank and gnomAD), and avoids the biases inherent in family studies. We identified 1,290 rare variants in 3,724 HCM cases (37%) and 604 rare variants in 756 DCM cases (34%). We report three sets of variants; i) 1015 variants found in cases but not identified in the population (caseAC>0, popAC=0; HCM n=740, DCM n=275), ii) 573 variants observed in a single case (caseAC=1, popAC>0; HCM n=265, DCM n=308), and most importantly (iii) 306 variants found multiple times in cases and controls (caseAC>1, popAC>0; HCM n=285, DCM n=21) for which a penetrance estimate can be computed. Reliable penetrance estimates for even a modest proportion of variants will be hugely impactful. For example, the most prevalent 50 variants observed recurrently in cases (caseAC>=10) account for about 1/3 of genetically explained cases (50 variants found in 1,182 individuals represent 32% of all G+ cases). The mean penetrance for these 50 variants is 35% (+/- 29% SD). This is the first dataset to report penetrance of individual recurrent variants at scale, and will inform the management of families with CM, and of individuals undergoing opportunistic genetic screening for secondary findings.

PrgmNr 2042 - Transferability of genetic loci and polygenic scores for cardiometabolic traits in ~22,000 British Pakistani and Bangladeshi individuals from a real-world healthcare cohort

View session detail

Author Block: **Q. Huang**[1], N. Sallah[2], D. Dunca[3], B. Trivedi[4], K. A. Hunt[5], S. Hodgson[6], S. Lambert[7], E. Arciero[8], Genes & Health Research team, C. Griffiths[4], J. Wright[9], R. C. Trembath[10], H. Hemingway[2], M. Inouye[7], S. Finer[5], D. A. van Heel[11], T. Lumbers[2], K. B. Kuchenbaecker[12], H. C. Martin[13]; [1]Wellcome Sanger Inst., Cambridge, United Kingdom, [2]Univ. Coll. London, London, United Kingdom, [3]UCL, London, United Kingdom, [4]Queen Mary Univ. of London, London, United Kingdom, [5]Blizard Inst., London, United Kingdom, [6]Univ. of Southampton, Southampton, United Kingdom, [7]Univ. of Cambridge, Cambridge, United Kingdom, [8]Wellcome Sanger Inst., Hinxton, United Kingdom, [9]Bradford Inst. for Hlth.Res., Bradford, United Kingdom, [10]King's Coll. London, London, United Kingdom, [11]Barts and The London, London, United Kingdom, [12]Univ Coll. London, London, United Kingdom, [13]Wellcome Trust Sanger Inst., Hinxton CB10 1SA, United Kingdom

Disclosure Block: **Q. Huang:** None.

South Asian-ancestry individuals have higher risk of heart disease than other ancestries. However, most genetic research has focused on European-ancestry individuals. It is largely unknown whether genetic loci and polygenic scores (PGS) associated with cardiometabolic traits are transferable from European-ancestry to British Pakistani and Bangladeshi individuals. Clinical utility of PGS has been shown in research settings. However, these findings may not generalise well to a real-world clinical setting, because many research cohorts are composed of volunteers who are healthier than the general population (e.g. UK Biobank), with clinical measures taken more comprehensively. The robustness of PGS applied to non-European individuals in a real-world healthcare system is largely unknown. To address these questions, we analysed data in Genes & Health (G&H), a British Pakistani and Bangladeshi cohort with linked electronic health records (N=~22,000). Two-thirds of people were recruited in healthcare settings. 97.4% of G&H participants are in the lowest two quintiles of the Index of Multiple Deprivation in the UK. We conducted genome-wide association studies (GWAS) of coronary artery disease (CAD), body mass index (BMI), lipid biomarkers and blood pressure in G&H. Trans-ancestry genetic correlations between UK Biobank Europeans and G&H for the tested traits were not significantly lower than 1 except for BMI which was nominally so (r_g=0.85, p=0.02). We observed high transferability of established loci with high power to replicate (â□□transferableâ□□) in G&H. We found evidence for sharing of causal variants between the two populations at 47-56% of the transferable loci associated with lipids and 26% of the transferable loci associated with BMI. PGS that were developed using data from primarily European-ancestry individuals showed variable transferability in G&H, with the relative accuracy compared to individuals of European ancestry (ratio of incremental R^2) â□¥95% for HDL-C, triglycerides, and blood pressure, but relatively lower for BMI (78%) and CAD (42%; ratio of incremental AUC). We observed significant improvement in categorical net reclassification in G&H (NRI=3.9%; 95% CI: 0.9-7.0%) when adding a previously developed CAD PGS to a clinical risk score (QRISK3). In conclusion, the genetic loci for CAD and its risk factors are largely transferable from European-ancestry studies to British Pakistanis and Bangladeshis, whereas the transferability of PGS varies greatly between traits. Our analyses indicate clinical validity for addition of PGS to existing clinical risk prediction tools in primary prevention settings for this population.

PrgmNr 2043 - Whole Genome Sequencing Association Study for Coronary Heart Disease in a Middle Eastern Cohort Validates Polygenic Risk Scores and Replicates Known Loci

View session detail

Author Block: **M. Saad**[1], K. Kunji[1], A. El-Menyar[2], E. Ullah[1], I. J. Kullo[3], J. Al Suwaidi[2]; [1]Qatar Computing Res. Inst., Hamad Bin Khalifa Univ., Doha, Qatar, [2]Hamad Med. Corp., Doha, Qatar, [3]Dept. of Cardiovascular Med., Mayo Clinic, Rochester, MN
Disclosure Block: **M. Saad:** None.
Background: The potential use of polygenic risk scores (PRSs) in clinical practice is tempered by concern about their portability among diverse populations. To prevent disparities in genomic medicine, there is an urgent need to conduct genome-wide association studies in non-European ancestry cohorts.
Methods: We conducted whole genome sequencing (WGS) with 30x coverage on coronary heart disease patients (n=1,067, mean age Â± SD = 59.96 years Â± 10.99; 70.32% males) and controls (n=6,170, mean age Â± SD = 40.02 years Â± 12.56 SD; 43.45% males) in a Middle Eastern cohort to compare the performance of available PRSs for CHD (LDpred, metaGRS, lassosum, and P+T) and identify common variants associated with CHD (via generalized linear mixed models).
Results: Excepting lassosum, all PRSs performed well. LDpred and metaGRS performed similarly (AUC= ~0.685) and outperformed P+T (AUC=0.667). Based on the OR per 1 SD increase (OR_{1sd}), P+T (OR_{1sd}=1.85 [1.69-2.02], P=3.69x10^{-41}) outperformed other PRSs (OR_{1sd}=1.61 [1.48-1.74], P=3.02x10^{-31} for LDpred and OR_{1sd}=1.61 [1.49-1.75], P=9.47x10^{-31} for metaGRS). After binning PRSs into 10 deciles, the odds of CHD in the top decile compared to all others was highest for metaGRS (3.87 [3.07-4.88]) and LDpred (3.45 [2.74-4.341]). Thirty-two known GWAS loci (e.g., *ABCG8*, *CELSR2*, and *SLC22A4*) were replicated in our analysis with *P*-5. Seven suggestive new loci/genes (*P*-6) with relevant biological function were identified (e.g., *CORO7*, *RBM47*, and *PDE4D*). The well-established 9p21 locus was not replicated.
Conclusions: Genome-wide PRSs derived from European ancestry cohorts performed well in a Middle Eastern cohort. Further studies are needed to develop and validate an ancestry specific PRS and to confirm the suggestive loci/genes.

PrgmNr 2044 - A whole genome sequencing study in a three-generation consanguineous family with morbid obesity, learning difficulty and failure of weight loss surgery, to identify susceptibility variants for obesity

View session detail

Author Block: **L. Al-Olabi**[1], N. Chami[2], S. K. Radha[1], Y. Wu[2], K. Musa[1], Y. Itan[2], A. J. Buckley[1], R. J. F. Loos[2], N. Lessan[1]; [1]Imperial Coll. London Diabetes Ctr., Abu Dhabi, United Arab Emirates, [2]The Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY
Disclosure Block: **L. Al-Olabi:** None.

Background

Family and twin studies show that genetic factors account for 40-70% of inter-individual variation in Body Mass Index (BMI). Monogenic obesity is rare, accounting for up to 6% prevalence in morbidly obese populations, with at least 10 genes currently identified. Within a population, genetic variation mediates inter-individual differences in susceptibility to the obesogenic environment and may also contribute to the inter-individual variation reported in weight loss outcomes following bariatric surgery.

Consanguinity has been key in the elucidation of many Mendelian and complex genetic diseases. Here, we combined the power conferred by consanguinity with genome sequencing for discovery of (rare) genetic variants in obesity.

Methods

We investigated a consanguineous Emirati family of 29 individuals with a history of morbid obesity. The proband had first-cousin parents and presented with morbid obesity, type 2 diabetes, with a history of gastric bypass which was later reversed and followed by a sleeve gastrectomy surgery years later. Onset of obesity in childhood was reported for the majority of individuals in the second and third generations. The proband was diagnosed with dyslexia, learning difficulty and Attention Deficit Hyperactivity Disorder (ADHD). Same features were self-reported in eight family members; one grandchild was diagnosed with ADHD. Nine of the eleven second generation siblings gave a history of at least one bariatric surgery; eight of these regained weight. Only one female sibling in the second generation, maintained a BMI within the normal range without surgical intervention. We looked for shared variants among individuals with the most severe obesity. Thus, analyses was performed in five individuals that still had obesity after at least one weight loss surgery, or those that did not have surgery but who have morbid obesity (BMI â¥40).

Results

The probandâ□□s karyotype was reported as normal. Array-based comparative genomic hybridization was negative and whole exome sequencing did not identify any causal variants in the proband. Thus, whole genome sequencing was carried out for 10 family members. Preliminary results from the genome analysis identified 25 rare nonsynonymous coding variants with a minor allele frequency

Conclusions

We have identified candidate variants including single nucleotide variations, deletions and insertions in genes, involved in signaling pathways that play an important role in processes linked to obesity. These results are currently being validated by Sanger sequencing.

PrgmNr 2045 - Deep shotgun metagenomic sequencing and functional profiling of the gut microbiota identifies associations with kidney function

View session detail

Author Block: **C. Nowak**[1], S. Sayols-Baixeras[2], G. Baldanzi[2], S. Ahmad[2], Y-T. Lin[2], K. F. Dekkers[2], GUTSY Research Collaboration, M. Orho-Melander[3], T. Fall[4], J. Ã□rnlÃ¶v[1,5,6]; [1]Karolinska Inst.t, Stockholm, Sweden, [2]Uppsala Univ., Uppsala, Sweden, [3]Lund Univ., MalmÃ¶, Sweden, [4]Uppsala Univ., Uppsala, Uppsala, Sweden, [5]Dalarna Univ., Falun, Sweden, [6]Region Dalarna, Falun, Sweden
Disclosure Block: **C. Nowak:** None.
Background: The human gut microbiota is composed of the bacteria, fungi and other microorganisms that live in the lower intestines in a symbiotic relationship with the host. Disruption of the gut microbiota has been associated with cardiovascular and metabolic diseases, but the association with kidney disease is still largely unknown. **Methods:** We studied the composition and predicted function of the gut microbiota based on shotgun whole-genome sequencing of microbial DNA in fecal samples collected from 9,788 adults enrolled in the longitudinal, population-based Swedish SCAPIS cohort study. Linear regression adjusted for technical variables, age, sex, Shannon diversity index and (in sensitivity analysis) established kidney disease risk factors was used to identify associations between the log(x+1)-transformed relative frequencies of 1,900 metagenomic species and estimated glomerular filtration rate (eGFR). The Benjamini-Hochberg false discovery rate (FDR) multiplicity correction was used. **Results:** We included 5,130 women (57.5Â±4.3 years) and 4,658 men (57.6Â±4.4 years). The mean eGFR was 86.5Â±11.3 for men, and 85.5Â±12.1 for women. Amongst all participants, 42% had an eGFR above 90, 39% had an eGFR between 75-90, 17% had an eGFR between 60-75, 2% had an eGFR between 45-60, and less than 0.1% had an eGFR below 45. In the age- and sex-adjusted model, we identified four bacteria that were associated with eGFR at an FDR Eubacteriales (two bacteria), *Coriobacteriales*, and *Veillonellales*. Gene set enrichment analysis indicated significant (FDR Conclusions: In the largest gut microbiome association study of kidney function in adults to date, we discovered four bacteria whose abundance was associated with glomerular filtration rate. The functional enrichment of kidney function-associated microbiota provides further insights into its possible role in kidney health.

PrgmNr 2046 - Estimation of variance explained by polygenic risk score of Crohnâ□□s disease and ulcerative colitis using Korean versus European GWAS

View session detail

Author Block: **S. Jung**[1], B. Ye[2], H-S. Lee[1], J. Baek[1], D. Park[1], Y. Kim[1], Y. Lee[1], S. Park[2], S-K. Yang[2], B. Han[3], J. Liu[4], K. Song[1]; [1]Dept. of Biochemistry and Molecular Biology, Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [2]Dept. of Gastroenterology, Asan Med. Ctr., Univ. of Ulsan Coll. of Med., Seoul, Korea, Republic of, [3]Dept. of Med., Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of, [4]Human Genetics Group, Genome Inst. of Singapore, Singapore, Singapore
Disclosure Block: **S. Jung:** None.
Recently polygenic risk score (PRS) attracted increasing interest from clinical community for their predictive value for multiple common diseases. The PRS estimates an individualâ□□s genetic liability to disease based on genotype profile and relevant genome-wide association study (GWAS) data. One of the most challenging aspects of moving PRS to the clinical use is ensuring that they are equally applicable to all health care users across ethnic groups. In this study, we compared the performance of the variance explained by PRS of Korean Crohnâ□□s disease (CD) and ulcerative colitis (UC) (725 CD, 1001 UC, and 378 controls) based on Korean (KOR) versus European (EUR) data using PRSice-2. The effect sizes were extracted from independent CD (KOR: 896 cases & 4,041 controls, EUR: 12,194 cases & 28,072 controls) and UC (KOR: 573 cases & 4,041 controls, EUR: 12,366 cases & 33,609 controls) GWAS. Despite the European studies had much larger samples size, PRS derived from Korean data (PRS_{KOR}) explained up to 14.3 % of phenotype variance of CD whereas those derived from European data (PRS_{EUR}) explained 9.9%. In contrast, for UC, the variance explained by PRS_{EUR} was far better than those explained by PRS_{KOR} (11.8% vs. 7.3%). After removing all SNPs in the *TNFSF15* and *HLA* region with the largest effect size in CD KOR GWAS, the variance explained by PRS_{KOR} became smaller than the variance explained by PRS_{EUR}. This suggested that the large variance explained by PRS_{KOR} might have been driven by these two loci with population-specific effects. The results highlight the need to perform large-scale genetic analyses in diverse populations and create tools for population genetic admixture to achieve an equitable benefit of PRS for IBD patients.

PrgmNr 2048 - HLA haplotype determines treatment response to GAD-alum immunotherapy in Type 1 diabetes: An applied example of a genetic precision medicine approach to pivotal clinical trial design

View session detail

Author Block: **U. Hannelius**[1], M. Widman[1], J. Ludvigsson[2]; [1]Diamyd Med., Stockholm, Sweden, [2]Div. of Pediatrics, Dept. of BioMed. and Clinical Sci., Faculty of Med. and Hlth.Sci. and Crown Princess Victoria Childrenâ□□s Hosp., LinkÃ¶ping Univ., LinkÃ¶ping, Sweden
Disclosure Block: **U. Hannelius:** Major Stockholder/Ownership Interest; Diamyd Medical. Salary/Employment; Diamyd Medical.
Background: Type 1 diabetes (T1D) results from an autoimmune destruction of insulin-producing beta cells. There is a great unmet need for disease-modifying treatments and antigen-specific immunotherapy using recombinant glutamate decarboxylase (GAD65) protein has shown mixed results across the broad population of T1D patients. Disease heterogeneity poses great challenges for clinical development in T1D, and a precision medicine approach targeting genetically defined subgroups can provide a solution. Our meta-analysis of 521 children and young adults (Hannelius et al. *Diabetologia* 2020, PMID: 32754804) with recently diagnosed T1D positive for GAD antibodies showed that recombinant GAD65 conjugated to aluminium hydroxide (GAD-alum/DiamydÂ®) had a significant and dose-dependent effect on the preservation of endogenous insulin secretion (stimulated C-peptide) over a 15 month period in patients carrying the HLA DR3-DQ2 haplotype, whilst no effect was seen in patients lacking DR3-DQ2. An even stronger effect was seen in patients with HLA DR3-DQ2 lacking the DR4-DQ8 haplotype. These HLA-specific effects were subsequently confirmed in a randomized double-blind placebo-controlled phase IIb trial (DIAGNODE-2, Ludvigsson et al. *Diabetes Care* 2021, PMID:34021020). Updated meta-analysis including the phase IIb data (n = 627) strengthened the results for C-peptide preservation and showed a significant effect on blood glucose control (reduction in HbA1c). **Implementation**: Diamyd Medical is launching a randomized double-blind placebo-controlled 24-month phase III trial (DIAGNODE-3) in the genetically defined population of recent-onset T1D patients carrying the HLA DR3-DQ2 haplotype (ca. 35% of all patients). Randomization will be stratified by the presence or absence of DR4-DQ8 (ca. 50% of DR3-DQ2-positive patients) and co-primary endpoints will be change in stimulated C-peptide and HbA1c. We will summarise the results from the meta-analyses and phase IIb trial and present details on the genetic precision-medicine strategy used in the design of the DIAGNODE-3 trial, including a broader rationale for the role of HLA subtypes in the response to antigen-specific treatment in T1D. **Conclusion**: We highlight the strategic and scientific rationale for the first pivotal phase III trial of a disease-modifying treatment for T1D based on a precision medicine approach applying a genetic marker to identify patients most likely to benefit.

PrgmNr 2049 - Selection pressures affecting Type 2 Diabetes in a South Indian and Scottish population

View session detail

Author Block: **C. Nangia**[1], S. Srinivasan[1], R. Venkatesan[2], V. Mohan[2], C. N. A. Palmer[1]; [1]Univ. of Dundee, Dundee, United Kingdom, [2]Madras Diabetes Res. Fndn., Chennai, India
Disclosure Block: **C. Nangia:** None.
Evolutionary forces have shaped how human populations respond to complex diseases. Type 2 Diabetes (T2D) in particular, has shown extensive genetic and phenotypic diversity worldwide. Selection studies can help identify the mechanisms underlying the diversity within and between populations. In this study, we aim to identify signatures of positive selection that have contributed to the phenotypic diversity between a Scottish and a novel South Indian T2D population.The study cohorts consisted of Scottish T2D individuals from the Genetics of Scottish Health Research Register (GoSHARE) study (n=6,681) and Genetics of Diabetes Audit and Research Tayside (Go-DARTS) study (n=9,487). South Indian T2D individuals were from the Madras Diabetes Research Foundation (MDRF)-cohort 1 (n=6,056) and cohort 2 (n=4,562).Whole genome pairwise F_{ST} was done on 12,737 T2D individuals from GoSHARE and MDRF cohort 1. High F_{ST} Single Nucleotide Polymorphisms (SNPs) were taken for further analysis. This was followed by tests for selective sweep - Tajimaâ□□s D and Nucleotide diversity. Candidate SNP association study with 13 phenotypes including Age at onset, anthropometry, blood pressure and lipids (age and sex adjusted; and sex stratified) was done in both the populations. Conditional analysis, fine mapping and gene expression studies were also done. A metanalysis of 4,855 high F_{ST} SNPs in 16,168 Scottish T2D individuals (GoDARTS and GoSHARE) showed the *ADAMTS9* locus to be significantly associated with obesity parameters - BMI and weight (*rs4422297,* prs16891982 (*SLC45A2*, pHLA locus was found to be significantly associated with age at onset, HDL and Triglycerides (*rs9273242, rs9273415* and *rs9273410* respectively, pSLC45A2 locus only in the Scottish population. No association was found in a metanalysis of the 2 Indian T2D cohorts. The results showed positive selection and association of the alleles in the Scottish cohort. BMI results favoured selection in the protective alleles indicating a non-thrifty genotype. The selective forces of these variants in the Scottish population do not appear to be due to an obesity related trait and may possibly represent pleiotropy due to other pressures such as exposure to infectious diseases. HLA results suggested immune-mediated beta cell dysfunction which may result in a less dyslipidemic form of diabetes. Absence of positive associations in the South Indian cohort was possibly due to gene-environment interactions.

PrgmNr 2050 - Validation and risk score applicability of T2D related variants in populations of East Asians and Europeans

View session detail

Author Block: **Y. Kim**, H-M. Jang, M. Hwang, B-J. Kim; Div. of Genome Sci., Dept. of Precision Med., NIH, Heungdeok-gu, Chungcheongbuk-do, Korea, Republic of
Disclosure Block: **Y. Kim:** None.
Genome-Wide Association Study (GWAS) enabled us to obtain unprecedented number of variants responsible for various diseases and related traits. With growing number of variants, polygenic risk score using discovered variants has gathered much attention in its application to clinical practice. However, majority of GWASs have been conducted in populations with European ancestry. These Euro-centric bias would lead a reduction in predicting diseases for non-Europeans. Therefore, discovered variants should be validated in trans-ethnic populations along with more GWASs in non-Europeans. In this study, we validated previously reported variants associated with type 2 diabetes (T2D) and related traits and its risk score applicability in East Asians (Korea Biobank Array(KBA), n=125,872) and Europeans from UK Biobank (UKB, n=337,475). As of Jan 2021, there were 8,823 variants associated with glycemic traits, lipids, liver enzymes, and T2D in GWAS catalog database. Based on the availability of imputed datasets in both KBA and UKB, 4,466 variants were selected for further analysis. The selected variants were further reduced to 2,913 variants considering linkage disequilibrium among the variants (r^2 50,000) implicating possible insufficient statistical power in this study. Risk score applicability was observed using T2D genetic risk score (GRS). T2D GRS was calculated using the selected variants and effect sizes from KBA, UKB, and trans-ethnic meta-analysis using KBA and UKB summary statistics. As expected, T2D-GRS using effect sizes from same population (KBA: 6.3%, UKB 5.1%) showed enhanced variance explained over T2D-GRS using effect sizes from different populations (KBA: 5.3%, UKB 4.1%). Also, trans-ethnic meta-analysis based T2D-GRS showed comparable performance to those of GRS using effect sizes from same population. This implicates trans-portability of T2D-GRS among populations. The validation results from our study provide valuable scientific evidence in using discovered variants and its risk score applicability across populations.

PrgmNr 2051 - Whole genome analysis of 342 adult cases of obesity and related metabolic disorders in highly consanguineous population

View session detail

Author Block: **M. Abedalthagafi**[1], H. Elbardis[1], F. Alqubaishi[1], A. AlMutairi[1], A. Alowayn[1], N. Aljohani[2], D. AlJaroudi[2], U. Thorsteinsdottir[3], K. Stefansson[4]; [1]King Abdulaziz City for Sci. and Technology, Riyadh, Saudi Arabia, [2]King Fahad Med. City, Riyadh, Saudi Arabia, [3]deCODE genetics, Reykjavik, IS, Iceland, [4]deCODE genetics, Reykjavik, Iceland
Disclosure Block: **M. Abedalthagafi:** None.
Background: Recent advantages allowing for direct sequencing of whole genomes or exomes offers the most comprehensive approach for extending discovery efforts to the detection rare sequence variants with large effects. Coupled with information about human disease and other traits, an unparalleled opportunity currently exists to identify rare coding variants with in genes that cause disease. Metabolic disorders like Obesity, Polycystic and type-2-diabetes (T2D) represent a worldwide epidemic that impose an enormous burden on public health. The Saudi population with one of the highest rate of consanguineous unions and high prevalence of both obesity and or T2D is ideal for identification, through whole-genome sequencing, of homozygous mutations with large effect on obesity and T2D. Method: A prospective study conducted at King Fahad Medical City (2019-2021) in Riyadh. All required ethical approval and consents obtained. Patients recruited from Obesity, and endocrinology clinics. DNA from 342 Saudi individuals (265 females and 83 males. Average age is 45-year-old. Average Body Mass Index (BMI) is 37.5) with high prevalence of obesity and/or T2D underwent whole-genome sequenced using illumine platform. 83 of the females also diagnosed with polycystic ovary syndrome (PCOS). Sequence variants called, annotated and screened for homozygous coding mutation that segregate with endocrinological disorders like obesity and/or T2D, PCOS within these patients. Result: Most samples show detectable inbreeding as expected. Average number of homozygous loss-of-function mutations per case is 0.9-1.2. Multiple known/expected pathogenic mutations identified in genes like *ABCC2, UPB1, HRG, FLT4, MSH3, TRAPPC2, CD36, CEL* and other. Likely pathogenic variants include: *PLIN1, LIPE, PAX4*. Homozygous/Hemizygous loss-of-function (LOF: essential splice, frameshift, stop gained) also reported as â□□Private â□□for our cohort in multiple novel genes (not in OMIM) like: *WDR54, ASB12, USP26, CTAG2, ZXDA*, and others. Conclusion: The Saudi population is one of the highest rate of consanguineous unions and high prevalence of both obesity and other related metabolic disorders. By analyzing whole-genome data from Saudis with obesity and/or T2D/ PCOS, we found homozygous mutations that identify novel genes that are causal in these conditions. Our data will provide insight into the pathophysiology of these conditions and potentially new targets for therapeutic developments.

PrgmNr 2052 - *HLA-A*11:01:01:01*, *HLA*C*12:02:02:01-HLA-B*52:01:02:02*, age and sex are associated with severity of Japanese COVID-19 with respiratory failure

View session detail

Author Block: **S-S. Khor**[1], Y. Omae[1], N. Nishida[2], M. Sugiyama[2], N. Kinoshita[3], T. Suzuki[3], M. Suzuki[3], S. Suzuki[4], S. Izumi[5], M. Hojo[5], N. Ohmagari[5], M. Mizokami[2], K. Tokunaga[6]; [1]Genome Med. Sci. Project, Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [2]Genome Med. Sci. Project, Natl. Ctr. for Global Hlth.and Med., Ichikawa, Chiba, Japan, [3]Disease Control and Prevention Ctr., Natl. Ctr. for Global Hlth.and Med. Hosp., Tokyo, Japan, [4]Biobank, Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [5]Dept. of Respiratory Med., Natl. Ctr. for Global Hlth.and Med. Hosp., Tokyo, Japan, [6]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan

Disclosure Block: **S. Khor:** None.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus causing coronavirus disease 2019 (COVID-19) was announced as an outbreak by the World Health Organization (WHO) in January 2020 and as a pandemic in March 2020. The majority of infected individuals have experienced no or only mild symptoms, ranging from fully asymptomatic cases to mild pneumonic disease. However, a minority of infected individuals develop severe respiratory symptoms. The objective of this study was to identify susceptible HLA alleles and clinical markers for the early identification of severe COVID-19 among hospitalized COVID-19 patients. A total of 137 patients with mild COVID-19 (mCOVID-19) and 53 patients with severe COVID-19 (sCOVID-19) were recruited from the Center Hospital of the National Center for Global Health and Medicine (NCGM), Tokyo, Japan for the period of February-August 2020. High-resolution sequencing-based typing for eight HLA genes was performed using next-generation sequencing. In the HLA association studies, HLA-A*11:01:01:01 [P_c = 0.013, OR = 2.26 (1.27-3.91)] and HLA-C*12:02:02:01-HLA-B*52:01:01:02 [P_c = 0.020, OR = 2.25 (1.24-3.92)] were found to be significantly associated with the severity of COVID-19. After multivariate analysis controlling for other confounding factors and comorbidities, HLA-A*11:01:01:01 [P = 3.34E-03, OR = 3.41 (1.50-7.73)], age at diagnosis [P = 1.29E-02, OR = 1.04 (1.01-1.07)] and sex at birth [P = 8.88E-03, OR = 2.92 (1.31-6.54)] remained significant. Early identification of potential sCOVID-19 could help clinicians prioritize medical utility and significantly decrease mortality from COVID-19.

PrgmNr 2053 - Applying the regional heritability mapping method to primary biliary cholangitis in the Japanese population

View session detail

Author Block: **O. Gervais**[1,2], K. Ueno[3], Y. Kawai[4], Y. Hitomi[5], Y. Aiba[6], M. Ueta[7], M. Nakamura[8,9,10], K. Tokunaga[4], M. Nagasaki[11]; [1]Nihon Univ., Mishima, Japan, [2]Human BioSci.s Unit for the Top Global Course CPIER, Kyoto Univ., Kyoto, Japan, [3]Genome Med. Sci. Project, Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [4]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [5]Hoshi Univ., Tokyo, Japan, [6]Clinical Res. Ctr., Natl. Hosp. Organization (NHO) Nagasaki Med. Ctr., Omura, Japan, [7]Kyoto Prefectural Univ. of Med., Kyoto, Japan, [8]Natl. Hosp. Organization (NHO) Nagasaki Med. Ctr., Omura, Japan, [9]Dept. of Hepatology, Nagasaki Univ. Graduate Sch. of BioMed. Sci., Omura, Japan, [10]Headquarters of PBC Res. in NHO Study Group for Liver Disease in Japan (NHOSLJ), Clinical Res. Ctr., NHO Nagasaki Med. Ctr., Omura, Japan, [11]Kyoto Univ., Kyoto-City, Japan

Disclosure Block: **O. Gervais:** None.

The development of GWAS in the early 2000s has contributed greatly to the discovery of susceptibility genes associated with complex traits. At the same time, the limitations of single-SNP GWAS have also created a need for other approaches, such as the regional heritability mapping (RHM) method, which makes use of multiple adjacent SNPs jointly to estimate the genetic effect of a given region of the genome. Although simulation studies have shown that RHM in many cases has higher detection power than single-SNP GWAS, it has been used so far primarily in agricultural research. In this study, we applied the RHM method to primary biliary cholangitis (PBC) in the Japanese population to assess its potential for the discovery of new susceptibility genes in human diseases. We identified three novel loci (*STAT4, ULK4,* and *KCNH5*) at the genome-wide significance level, two of which (*ULK4* and *KCNH5*) have not been found associated with PBC in any population previously. Importantly, these three genes could not be detected by using conventional single-SNP GWAS. Our results underscore the potential of the RHM method for the discovery of new loci in human diseases, and provide strong empirical evidence that RHM is a useful tool that can be used as an effective complement to single-SNP GWAS. In addition, we conducted an analysis of liver tissue mRNA microarray data, which revealed higher gene expression levels in *ULK4* in PBC patients (P

PrgmNr 2054 - Common, low frequency, rare and ultra-rare variants contribute to COVID-19 severity

View session detail

Author Block: **A. Renieri**[1,2,3], N. Picchiotti[4,5], K. Zguro[2], C. Fallerini[1,2], S. Daga[1,2], M. Baldassarri[1,2], F. Fava[1,2,3], M. Tanfoni[4], E. Benetti[2], F. Valentino[1,2], G. Doddato[1,2], A. Giliberti[1,2], R. Tita[3], S. Amitrano[3], M. Bruttini[1,2,3], L. Di Sarno[1,2], D. Alaverdian[1,2], G. Beligni[1,2], S. Croci[1,2], M. Palmieri[1,2], I. Meloni[1,2], A. Pinto[3], C. Gabbi[6], S. Ceri[7], A. Esposito[7], M. Gori[4,8], E. Frullanti[1,2], F. Mari[1,2,3], GEN-COVID Multicenter Study, S. Furini[2]; [1]Med. genetics, Univ. of Siena, Siena, Italy, [2]Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies, Univ. of Siena, Siena, Italy, [3]Genetica Medica, Azienda Ospedaliero-Univ.ria Senese, Siena, Italy, [4]Univ. of Siena, DIISM-SAILAB, Siena, Italy, [5]Dept. of Mathematics, Univ. of Pavia, Pavia, Italy, [6]Independent Med. Scientist, Milano, Italy, [7]Politecnico di Milano, DEIB, Milano, Italy, [8]UniversitÃ© CÃ´te dâ□□Azur, Inria, CNRS, I3S, Maasai, Bordeaux, France

Disclosure Block: **A. Renieri:** None.

Host genetics is an emerging theme in COVID-19. A handful of common polymorphisms and some rare variants have been identified, either through GWAS or candidate gene approach. However, an organic model is still missing. We applied a gene based approach to Whole Exome Sequencing data of 2,200 SARS-CoV-2 infected subjects with different outcomes, from very severe to oligo-asymptomatic, within GEN-COVID cohort. Common, low frequency, rare and ultra-rare variants were tested separately. Autosomal genes were tested separately from genes on the X chromosome. LASSO logistic regression was applied under the assumption of either dominant or recessive phenotype. Both sexes were tested together and separately. Among 18,347 tested genes, about 2,000 genes were found to be involved after multiple train-test splits. Among these, Â¼ of genes was sex specific. These latter genes are under sexual hormone control, such as testosterone down-regulated TLR3 gene or estrogen down-regulated TLR5 gene, and exert their effect in one sex only. Integrated PolyGenic Score (IPGS) was calculated as ($n_{mildness}$ - $n_{severity}$ common) + F1 ($m_{mildness}$ - $m_{severity}$ low-frequency) + F2 ($x_{mildness}$ - $x_{severity}$ rare) + F3 ($y_{mildness}$ - $y_{severity}$ ultra-rare), where F is a weighting factor taking into account that the impact of protein function inversely correlates with frequency. The weighting factor for ultra-rare is so high that, when the ultra-rare is present, the model simulates Mendelian inheritance such that it happens in TLR7 ultra-rare variants, which we and others proved to be an inherited X-linked form of COVID-19. IPGS is able to improve prediction of clinical outcome in addition to the already known powerful factors, such as age and comorbidities, and profiles a specific genetic signature for adjuvant therapy. A platform of clinical trials based on genetic markers has been submitted to Italian Drug Agency AIFA. This is the first modelling of a genetic complex disease taking into account whole variability from common to ultra-rare variant and it can be translated to many other complex disorders.

PrgmNr 2055 - Comprehensive analysis of HLA association in Japanese with childhood-onset nephrotic syndrome

View session detail

Author Block: **X. Jia**[1], S-S. Khor[2], Y. Kawai[3], T. Horinouchi[4], K. Iijima[5], K. Tokunaga[3]; [1]Natl. center for global Hlth. and medicine, Tokyo, Japan, [2]Natl. Ctr. for Global Hlth.and Med., Tokyo, Tokyo, Japan, [3]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [4]Dept. of Pediatrics, Kobe Univ. Graduate Sch. of Med., Kobe, Japan, [5]Hyogo Prefectural Kobe Childrenâ□□s Hosp., Kobe, Japan
Disclosure Block: **X. Jia:** None.
Background: Idiopathic nephrotic syndrome (INS) is the most common kidney disease in children. Corticosteroids are used as the first-line treatment and are effective in around 80% pediatric patients (steroid-sensitive nephrotic syndrome, SSNS). About 10-20% patients showed resistant to the steroid therapy and have poor outcome, which are named as steroid-resistant nephrotic syndrome (SRNS). *HLA-DR/DQ* region has been highlighted as the most important genetic factor by previous genome wide association studies (GWASs) in multiple populations. Methods: To obtain a comprehensive understanding of *HLA* association with childhood INS, pediatric patients with SSNS (N=1,088), SRNS (N=266) and healthy adults (N=3,331) with Japanese ancestry were recruited. Samples were genotyped by Japanese-specific Affymetrix â□□Japonica Arrayâ□□. Fourteen HLA genes (*HLA-A, -C, -B, -DRB1, -DQB1, -DPB1, -DRA, -DQA1, -DPA1, -DPA2, -DMA, -DMB, -DOA* and *-DOB*) were imputed at 3-field level using Japanese-specific reference panel by â□□HIBAGâ□□. Results: In total, 838 SSNS patients, 220 SRNS patients and 2,822 controls passed the post quality control using a cut-off threshold (CT) >= 0.4 for the probability of imputed HLA genes. *HLA-DRB1*08:02:01-DQB1*03:02:01* was the most significant risk factor associated with Japanese childhood INS (SSNS vs. controls: OR=3.31 [2.55-4.29], P-corrected=1.33E-21; SRNS vs. controls: OR=6.29 [4.46-8.79], P-corrected=9.22E-35). *HLA-DQA1*01:02:01* was identified as the most significant protective allele for childhood INS (SSNS vs. controls: P-corrected=2.01E-28, odds ratio [OR]=0.28 [0.22-0.35]; SRNS vs. controls: P-corrected=7.36E-10, OR=0.22 [0.13-0.36]). *HLA-DRB1*08:02:01-DQB1*03:02:01* showed an increased frequency in SRNS patients than SSNS patients (13.2% in SRNS while 7.4% in SSNS, P-corrected=2.24E-03, OR=1.90 [1.34-2.67]). Conclusions: We clarified *HLA* factors associated with childhood INS and the responsiveness to steroid therapy in the Japanese population using high resolution HLA-imputation, providing deeper insight into the disease mechanism and useful biomarker for predicting patientsâ□□ responsiveness to steroid treatment.

PrgmNr 2056 - Effective way that determining mild or severe COVID-19 patient only using T-cell receptor(TCR) sequencing data

View session detail

Author Block: **J. Han**; Seoul Natl. Univ., Dept. of BioMed. Sci., Seoul, Korea, Republic of
Disclosure Block: **J. Han:** None.
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) led to the coronavirus disease 2019 (COVID-19) pandemic situation. Furthermore, until now, people do not fully understand SARS-CoV-2 even though there are already a lot of mild or severe patient sequencing data who suffer from COVID-19. SARS-CoV-2 have ignored peopleâ□□s immune system including T cell activity, easily infecting many people, and at last have arisen the pandemic situation. Therefore, analysing these dozens of patients T-cell receptor(TCR) sequencing data is seriously important because it would give us insight into researching SARS-CoV-2 mutation phenotypes and preparing other severe viruses after COVID-19 situation. Here, we compare TCR clustering toolsâ□□ accuracy to which tool most nearly puts T-cells together normal control, mild or severe patientâ□□s groups by only using subject TCR sequencing data.

PrgmNr 2057 - Fine-mapping of novel susceptibility loci associated with eosinophil granule proteins (ECP and EDN) reveals five biologically relevant genes for asthma

View session detail

Author Block: **R. Vernet**[1], R. Matran[2], F. Zerimech[3], A-M. Madore[4], P. Margaritte-Jeannin[1], M-H. Dizier[1], F. Demenais[1], C. Laprise[4], R. Nadif[5], E. Bouzigon[1]; [1]UniversitÃ© de Paris, INSERM UMR 1124, Group of Genomic Epidemiology of Multifactorial Diseases, Paris, France, [2]UniversitÃ© Lille and CHU de Lille, Lille, France, [3]PÃ´le de Biologie Pathologie GÃ©nÃ©tique, Laboratoire de Biochimie et Biologie MolÃ©culaire, CHU de Lille, Lille, France, [4]Basic Sci. Dept., UniversitÃ© du QuÃ©bec Ã Chicoutimi, Saguenay, QuÃ©bec, Canada, Ctr. intersectoriel en santÃ© durable, UniversitÃ© du QuÃ©bec Ã Chicoutimi, Saguenay, QC, Canada, [5]UniversitÃ© Paris-Saclay, UVSQ, Univ. Paris-Sud, Inserm, Equipe dâEpidÃ©miologie Respiratoire IntÃ©grative, CESP, Villejuif, France
Disclosure Block: **R. Vernet:** None.
Background: Eosinophils play a key role in the allergic response in asthma by the release of cytotoxic molecules such as eosinophil cationic protein (ECP) and eosinophil-derived neurotoxin (EDN) that generate epithelium damages.
Objective: We aimed to identify genetic variants influencing ECP and EDN levels in asthma-ascertained families of European ancestry.
Methods: We conducted univariate and bivariate genome-wide association analyses of these proteins in 1,018 subjects from the EGEA study with follow-up in 153 subjects from SLSJ study and performed meta-analysis to combine evidence from the two datasets. We then conducted Bayesian statistical fine-mapping together with in silico quantitative trait locus and functional annotation analyses to identify credible sets of variants and target candidate genes.
Results: We identified four genome-wide significant loci (*P*-8) including six distinct signals associated with ECP and/or EDN levels. These six signals were located on 14q11, 7p21, 1p31 and 9q22 chromosomal regions. Four of the six distinct signals were fine-mapped to small credible sets of putative causal variants including at most 10 SNPs with sum of posterior inclusion probability (PIP) of at least 95%. Moreover, two of these credible sets related to 7p21 region included a SNP with high PIP (greater than 0.7). Functional annotation analyses and the literature showed that genes targeted by these SNPs have biological relevance related to the pathophysiological mechanisms of eosinophil activity and asthma: *RNASE2* and *RNASE3* (14q11) encode EDN and ECP respectively; *AK4* (1p31) codes for an adenylate kinase that has been shown to regulate several inflammatory genes; *NDUFA4* (7p21) codes for a component of the mitochondrial respiratory chain; and *CTSL* (9q22) encodes a lysosomal cysteine proteinase involved in the immune response and has previously been identified as part of a gene module associated with childhood-onset asthma.
Conclusion: This study highlights the interest of joint analysis of biological phenotypes involved in the pathophysiological mechanisms of asthma to increase power to detect new loci and potentially functional genes.
Funded: AAP Nord-Pas-de-Calais. ANR-GenCAST, IRSC

PrgmNr 2058 - Genetic diversity and epidemiology of human rhinovirus among children with severe acute respiratory tract infection in Guangzhou of China

View session detail

Author Block: **H. LI**, W. LI, B. Xue, Z. Deng; Jinan Univ., Guangzhou, China
Disclosure Block: **H. Li:** None.
Human rhinovirus (HRV) is one of the major viruses of acute respiratory tract disease among infants and young children. HRV is increasingly recognized not only as a cause of mild upper respiratory tract infection, but also in more severe lower respiratory tract infections, such as pneumonia, bronchiolitis and asthma. HRVs, members of the family *Picornaviridae* and the genus *Enterovirus*, are positive-sense, single-stranded-RNA (ssRNA) viruses of approximately 7,200 bp. The viral genome consists of a single gene whose translated protein is cleaved by virally encoded proteases to produce 11 proteins. HRVs were classified into three species, HRV-A, -B, and-C based on phylogenetic sequence criteria. Nasopharyngeal swabs from a total of 655 children with acute respiratory illness in the Guangdong Maternal and Child Health Hospital and Huaqiao Hospital were obtained from August 2018 to December 2019. HRV was screened for by a real-time reverse-transcription PCR targeting the viral 5â□²UTR. HRV was detected in 6.41% of the 655 specimens. Among these 40 positive samples, HRV infection was frequently observed in children under 2 years old (57.13%). HRV-A and HRV-C were detected in 18 (45%) and 22 (55%) specimens. Children who experienced rhinorrhoea were more common in the HRV-C infection patients than HRV-A. The viral load was higher in HRV-C detection group than HRV-A detection group (p = 0.0148). The median peak symptom score was higher in patients with HRV-C infection as compared to HRV-A (p = 0.0543), even though the difference did not reach significance. Our findings expand knowledge of HRV infectionsâ□□ clinical spectrum.

PrgmNr 2059 - Genetic Similarity Assessment of Qatar and Italy Populations in the context of COVID-19

View session detail

Author Block: **H. Mbarek**[1], M. Cocca[2], Y. Al-Sarraj[1], C. Saad[1], W. AlMuftah[1], R. M. Badji[1], A. Novelli[3], A. Al-Tawashi[4], S. Salvaggio[4], A. Althani[1], G. Novelli[5], S. I. Ismail[1]; [1]Qatar Fndn., Qatar Sci. & Technology Park Innovation Ctr., room 356, Doha, Qatar, [2]Univ. of Trieste, Trieste, Italy, [3]Bambino GesÃ¹ Hosp.,, Rome, Italy, [4]Qatar Fndn. Res., Dev. and Innovation, Qatar Fndn., Doha, Qatar., Doha, Qatar, [5]Dept. of Biomedicine and Prevention, Tor Vergata Univ. of Rome, Rome, Italy, Rome, Italy
Disclosure Block: **H. Mbarek:** None.
COVID-19 continues to spread worldwide with over three million deaths to date and rising. However, this global spread is coupled with stark anomalies in morbidity and mortality. These differences can be observed not only between different populations but also within the same population. Host genomic information, specifically genomic variations, may characterize susceptibility to disease and identify people with a higher risk of harm, leading to a better targeting of care and vaccination. Italy was the epicenter for the spread of COVID-19 in Europe, the first country to go into a national lockdown and has one of the highest COVID-19 associated mortality rates. Qatar on the other hand, despite having one of the highest worldwide number of laboratory-confirmed cases, has a very low mortality rate.In this study we compared whole-genome sequencing data of 14398 adults and Qatari-national, to 925 Italian individuals. We also included in the comparison whole-exome sequence data from 196 Italian laboratory confirmed COVID-19 cases. We focused our study on a curated list of 3619 candidate genes involved in innate immunity and host-pathogen interaction. Two population-gene metric scores, the Delta Singleton-Cohort variant score (DSC) and Sum Singleton-Cohort variant score (SSC), were applied to estimate the presence of selective constraints in the Qatari population as well as the Italian cohorts.Results based on DSC & SSC metrics demonstrated a different selective pressure on some gene between Qatari and Italian populations. Example of these genes include TLR3 and IFNAR1 which are already reported by other recent studies to be involved in life-threatening COVID-19 cases.This study highlighted the genetic differences between Qatari and Italian populations and identified a subset of genes involved in innate immunity and host pathogen interaction.

PrgmNr 2060 - High-resolution genomic architecture of COVID-19 severe disease using multi-ethnic whole genome sequencing data

View session detail

Author Block: **A. Kousathanas**[1], E. Pairo-Castineira[2,3], A. Stuckey[1], C. Odhams[1], S. Walker[1], S. Clohisey[2], K. Rawlik[2], A. Fawkes[4], D. J. Rhodes[1], A. Siddiq[1], P. Goddard[1], S. Donovan[1], ISARIC4C Investigators, Genomics England Research Consortium, GenOMICC Investigators, C. P. Ponting[3], K. Rowan[5], L. Murphy[4], P. J.M. Openshaw[6], M. G. Semple[7], A. Rendon[1], R. H. Scott[1], A. Law[2], L. Moutsianas[1], M. Caulfield[1,8], K. J. Baillie[2,3,9]; [1]Genomics England, London, United Kingdom, [2]Roslin Inst., Univ. of Edinburgh, Edinburgh, United Kingdom, [3]MRC Human Genetics Unit, The MRC Inst. of Genetics and Molecular Med., Univ. of Edinburgh, Western Gen. Hosp., Edinburgh, United Kingdom, [4]Edinburgh Clinical Res. Facility, Western Gen. Hosp., Univ. of Edinburgh, Edinburgh, United Kingdom, [5]Intensive Care Natl. Audit & Res. Ctr., London, United Kingdom, [6]Natl. Heart and Lung Inst., Imperial Coll. London, London, United Kingdom, [7]NIHR Hlth.Protection Res. Unit for Emerging and Zoonotic Infections, Inst. of Infection, Vet. and Ecological Sci. Univ. of Liverpool, Liverpool, United Kingdom, [8]William Harvey Res. Inst., Barts and the London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom, [9]Intensive Care Unit, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom

Disclosure Block: **A. Kousathanas:** None.

With the aim of exploring host genetic factors underlying COVID-19 severity, the Genetics Of Mortality In Critical Care consortium (GenOMICC) and Genomics England are recruiting participants with severe and mild COVID-19. Our previous work in the GenOMICC consortium revealed therapeutically-targetable variants underlying life threatening COVID-19 (Pairo-Castineira et al. 2020). Working with the global Host Genetics Initiative consortium (HGI), we have shown that the strongest genetic signals are present in the critically ill population. However, most previous work, including our own, has used genotyping arrays and largely focused on a subset of common genetic variation. In order to facilitate a high-resolution analysis of the genome and exploration of all types of genetic variation underlying disease risk, we are performing whole-genome sequencing (WGS) of up to 35,000 participants with severe and mild COVID-19. To increase statistical power, we are combining these data with WGS data generated through the 100,000 Genomes Project from a wide range of ancestral backgrounds. Recruitment is ongoing with successive rounds of analysis and data freezes being made available in a cloud-based research environment to facilitate international research efforts. Our latest data freeze consists of over 10,000 individuals with COVID-19, including the largest WGS cohort of critically ill COVID-19 patients assessed to-date. GWAS analysis of data on a freeze of 4,677 patients, replicated genetic associations previously identified by us, HGI and other groups on chromosomes 3,12,19, and 21. The WGS data enabled precise fine-mapping that narrowed down the 3p21.31 association cluster to two independent signals. Trans-ancestry GWAS analysis showed that one of the association signals on 3p21.31 can be independently replicated in multiple ethnicities but with significantly different effect sizes, underscoring the heterogenous impact of genetic ancestry on COVID-19. Rare variant burden analyses have yet to identify robustly any genes associated with severe COVID-19, which is consistent with other recent findings (Kosmicki et al. 2021). Our study provides the first comprehensive investigation of the contribution of rare and common genetic variation on COVID-19 disease outcomes using large scale WGS data, which will be instrumental in elucidating the genomic architecture of the disease and inform therapeutic interventions.

PrgmNr 2061 - Identification of causal genetic variants in invasive pneumococcal diseases by exome analysis in children

View session detail

Author Block: **M. GÃ©lin**[1], S. Limou[1], P-A. Gourraud[1], A. Durand[1], O. Rousseau[1], C. Gras-Leguen[2], F. Lorton[2], E. Launay[2], N. Vince[2]; [1]CRTI, UniversitÃ© de Nantes, Nantes, France, [2]UniversitÃ© de Nantes, Nantes, France

Disclosure Block: **M. GÃ©lin:** None.

Invasive pneumococcal diseases (IPDs) are still severe diseases in the pediatric population, despite major progress in vaccination and therapeutics. Apart from the pathogen virulence factors, host individual characteristics, including host genetic factors, are implicated in the susceptibility and severity to these diseases. Further understanding of the molecular factors involved in the development of these severe infections could lead to better treatment, prevention, and thus better outcome. Previous studies identified several candidate genes for IPDs, however, mostly because of methodological shortcomings, these results are yet to be confirmed. Our study aimed to identify rare coding genetic variants implicated in the development of IPDs through a large-scale genetic strategy using whole-exome sequencing (WES) of 32 children admitted in the pediatric intensive care unit for IPD. Contrary to most previous studies related to IPDs, we chose an untargeted approach to identify novel variants. We developed a bioinformatic analysis pipeline to identify rare, non-synonymous and possibly pathogenic variants implicated in IPDs. The pipeline was designed to identify genetic variants and annotate their Minor Allele Frequency (MAF) and their possible pathogenicity according to several tools: VEP, ANNOVAR, InterVar and CADD. Furthermore, we compared the variant distribution observed in our patients to an unrelated control population (n = 69) in order to validate our pipeline and discoveries. We identified 86 rare variants at the heterozygous state in 18 different genes, including 7 genes previously associated to immunological functions, inflammation, or infectious diseases: *AHNAK2, AK2, CFTR, CYP4F2, ESRRA, FCGBP* and *PABPC1.* The control population presented significantly fewer variants ($p=7.6x10^{-24}$) in fewer genes ($p=7.8\ x10^{-27}$): 5.8 variants in 3.1 genes / individual *vs* 14.9 variants in 10.5 genes / individual in the IPD population. Our results have therefore revealed multiple heterozygous variants in a restricted set of genes for each patient presenting an extreme phenotype of pneumococcal disease, emphasizing the likely polygenicity of IPD risk. We will pursue our effort in better characterizing these associations using *burden test* strategies. In conclusion, these first results suggest an unprecedented immune deficiency involving several rare mutations in immune-related genes.

PrgmNr 2062 - Identifying novel causative mutations forDOCK8 immunodeficiency syndromeusing whole exome sequencing

View session detail

Author Block: **R. Noroozi**[1,2]; [1]Jagiellonian Univ., Krakow, Poland, [2]Malopolska Ctr. of Biotechnology, Jagiellonian Univ., Krakow, Poland
Disclosure Block: **R. Noroozi:** None.
DOCK8 immunodeficiency syndrome (DIDS) is a rare autosomal recessive (AR) disorder characterized by elevated serum IgE levels, eosinophilia, recurrent cutaneous infections, severe eczema, and sinopulmonary and gastrointestinal infections. This syndrome is a multisystem disease that is associated with both immune deficiency and neurological complications. In this study, we describe the clinical characteristics of two Iranian patients with DOCK8 deficiency and propose possible mechanisms for this condition. By using whole-exome sequencing (WES), we identified two novel mutations, namely c.3233_3234del AG (p.Q1078fs) in exon 6 and a large deletion with 94 kb (c.405-3231 deletion, p.K135fs), in these two patients. These variations are confirmed with Sanger sequencing and CGH array. The subsequent co-segregation analysis is performed to identify inheritance patterns. Both patients were homozygote, and their parents were heterozygote for the variations. For further investigation, prediction tools were applied to identify the pathogenicity of the variations and for modeling the truncated proteins. The patients did not show neurological abnormalities associated with a deficiency of the N terminal region of DOCK8. The absence of neurological complications in the first patient is justifiable due to the maintenance of the proline-rich region in DOCK8, but for the second patient with expanded deletion which is almost like null DOCK8 protein, it is not presumable, pointing to the fact that the C terminal region of the protein might have functions in the proliferation and migration neurons in the peripheral nervous system. Alternatively, neurological abnormalities may follow an age-dependent pattern, leading to the appearance of related symptoms later in life. Further multiple functional studies are needed to model different identified variants in animal models to confirm our results and suggest possible mechanisms associated with DOCK8 deficiency in this study.

PrgmNr 2063 - PADI4 and PADI2 enhance collagen-initiated inflammatory responses

View session detail

Author Block: **A. Suzuki**[1], T. Shibuya[1], K. Yamamoto[2]; [1]RIKEN, Yokohama, Japan, [2]RIKEN Ctr. for Integrative Med. Sci., Yokohama, Kanagawa, Japan
Disclosure Block: **A. Suzuki:** None.
Previously, peptidylarginine deiminase type 4 (PADI4) was identified as a susceptibility gene for Rheumatoid arthritis (RA) by genome-wide association studies. Peptidyl citrulline is a target antigen of anti-citrullinated peptide antibodies (ACPAs), and only PADs (translated protein from PADI genes) can provide peptidyl citrulline via modification of protein substrates. Also the distribution of PADI4 and PADI2 has overlap in immune cells. The aim of this study was to investigate the relationship between PADI4 gene and PADI2 gene in the progression of RA. To clarify the physiological function of PADI4 and PADI2 in RA, we used collagen-induced arthritis (CIA), known as a RA model mouse. We examined that localization of PAD4 and PAD2 protein was indicated by immunohistochemistry in CIA mice. We also measured expression of Padi genes and various inflammatory cytokines in immune cells by real-time TaqMan assay and ELISA, respectively. We generated PADI4â□□/â□□ and PADI2â□□/â□□ mice and performed experimental arthritis. We demonstrated that the clinical disease score was significantly decreased in PADI4â□□/â□□ mice and PADI4 expression was induced by CII immunization. In PADI4â□□/â□□ mice sera, serum anti-type II collagen (CII) IgM, IgG, and inflammatory cytokine levels were also significantly decreased compared with those in wild-type mice sera. Interestingly, PADI2 expression was compensationally induced in CD11b+ cells of PADI4-/- mice. Furthermore, we examined that the clinical disease score of CIA mimce and expression levels of Padi genes in PADI2â□□/â□□ CIA mice. It appeared that collagen-initiated inflammatory responses were reduced in PADI2â□□/â□□ CIA mice. However, gene expression levels of CIA of Padi2 using wild type mice was not observed significant difference between CIA and control mice. As the one of the reason, it is possible that the loss of function of the Padi2 gene may affect the reduction of the arthritis score not due to gene expression levels.
This study was supported by Grants-in-Aid for Scientific Research (C).

PrgmNr 2064 - rs1944919 on human chromosome 11q23.1 and its effector genes *COLCA1* and *COLCA2* confer susceptibility to primary biliary cholangitis

View session detail

Author Block: **Y. Hitomi**[1], Y. Aiba[2], Y. Kawai[3], K. Kojima[4], K. Ueno[5], N. Nishida[6], M. Kawashima[7], O. Gervais[8], S-S. Khor[9], M. Nagasaki[10], K. Tokunaga[3], M. Nakamura[11], M. Tsuiji[1]; [1]Hoshi Univ., Tokyo, Japan, [2]Natl. Hosp. Organization (NHO) Nagasaki Med. Ctr., Omura, Japan, [3]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [4]Tohoku Univ, Sendai, Japan, [5]Genome Med. Sci. Project, Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [6]Natl. Ctr. for Global Hlth.and Med., Ichikawa, Chiba, Japan, [7]Japan Sci. and Technology Agency, Chiyoda-ku, Japan, [8]Nihon Univ., Mishima, Japan, [9]Natl. Ctr. for Global Hlth.and Med., Tokyo, Tokyo, Japan, [10]Kyoto Univ., Kyoto-City, Japan, [11]Natl. Hosp. Organization (NHO) Nagasaki Med. Ctr., Omura, Nagasaki, Japan

Disclosure Block: **Y. Hitomi:** None.

Primary biliary cholangitis (PBC) is a chronic and cholestatic autoimmune liver disease caused by the destruction of intrahepatic small bile ducts. The higher monozygotic/dizygotic (MZ/DZ) ratio and the higher estimated relative sibling risk (Î»s) in PBC patients as compared to unaffected individuals indicates the involvement of strong genetic factors in the development of PBC. Our previous genome-wide association study (GWAS) identified chromosome 11q23.1 (including *POU2AF1*, *COLCA1*, and *COLCA2*) as the susceptibility locus for PBC in the Japanese population. Although candidate genes with well-known functions that are located in susceptibility loci (i.e. *POU2AF1*) have often been reported as disease susceptibility genes in GWASs, the identification of effector genes regulated by primary functional variants is necessary to understand the contribution of susceptibility loci to pathogenesis. Here, in order to analyze the disease susceptibility of all genetic variations, we carried out high-density association mapping of chromosome 11q23.1 based on single nucleotide polymorphisms (SNPs) imputation using data from a whole-genome sequence reference panel of 1,070 Japanese individuals and our previous GWAS. Subsequent *in silico* and *in vitro* functional analyses identified rs1944919 as the primary functional variant. Expression-quantitative trait loci (e-QTL) analyses in the esophageal mucosa showed that rs1944919 was significantly associated with expression levels of *COLCA1 and COLCA2* (P = 3.8 Ã□ 10^{-29} and P = 6.5 Ã□ 10^{-32}, respectively) whose function have not been fully elucidated. Concordant association was detected in e-QTL analysis in B-cell (P = 0.00023 and P = 0.00040, respectively). Despite a distance of 100 kb between the *COLCA* genes and rs1944919, chromatin interactions between rs1944919 and upstream sequences of the *COLCA1/COLCA2* was detected in GM12878 cells. Additionally, the effects of rs1944919 on *COLCA1/COLCA2* expression levels were confirmed using genotype knock-in versions of Raji (human B lymphocytes) and Jurkat (human T lymphocytes) constructed using the CRISPR/Cas9 system and differed between rs1944919-G/G clones and -T/T clones (P COLCA1 and *COLCA2* are the effector gene regulated by the primary functional SNP rs1944919, and that increased expressions of *COLCA1/COLCA2* might be involved in the pathogenesis of PBC.

PrgmNr 2065 - The impact of SARS-CoV-2 multiple spike protein variants on COVID-19 outcomes

View session detail

Author Block: **.. Gunadi**[1], H. Wibawa[2], M. Hakim[3], .. Marcellus[1], I. Trisnawati[4], R. Khair[5], K. Iskandar[6], .. Siswanto[7], E. Supriyati[8], D. Nugrahaningsih[9]; [1]Genetics Working Group/Pediatric Surgery Div., Dept. of Surgery, Faculty of Med., Publ. Hlth.and Nursing, Univ.s Gadjah Mada/Dr. Sardjito Hosp., Yogyakarta, Indonesia, [2]Disease Investigation Ctr. Wates, Directorate Gen. of Livestock and Animal Hlth.Services, Ministry of Agriculture, Yogyakarta, Indonesia, [3]Dept. of Microbiol., Faculty of Med., Publ. Hlth.and Nursing, Univ.s Gadjah Mada/Dr. Sardjito Hosp., Yogyakarta, Indonesia, [4]Pulmonology Div., Dept. of Internal Med., Faculty of Med., Publ. Hlth.and Nursing, Univ.s Gadjah Mada/Dr. Sardjito Hosp., Yogyakarta, Indonesia, [5]Dept. of Clinical Pathology and Lab. Med., Faculty of Med., Publ. Hlth.and Nursing, Univ.s Gadjah Mada/Dr. Sardjito Hosp., Yogyakarta, Indonesia, [6]Dept. of Child Hlth./Genetics Working Group, Faculty of Med., Publ. Hlth.and Nursing, Univ.s Gadjah Mada/ UGM Academic Hosp., Yogyakarta, Indonesia, [7]Dept. of Physiology, Faculty of Med., Publ. Hlth.and Nursing, Univ.s Gadjah Mada/UGM Academic Hosp., Yogyakarta, Indonesia, Yogyakarta, Indonesia, [8]Ctr. of Tropical Med., Faculty of Med., Publ. Hlth.and Nursing, Univ.s Gadjah Mada, Yogyakarta, Indonesia, [9]Dept. of Pharmacology and Therapy/Genetics Working Group, Faculty of Med., Publ. Hlth.and Nursing, Univ.s Gadjah Mada, Yogyakarta, Indonesia

Disclosure Block: **.. Gunadi:** None.

Background: Recent studies on the association of SARS-CoV-2 variants with COVID-19 outcomes have been reported, focusing on variant of concerns (VOC). However, studies on the impact of other mutations within spike (S) protein of SARS-CoV-2 on COVID-19 illness are very limited. Moreover, virulence of SARS-CoV-2 might be affected by geographic condition. We determined: 1) association between SARS-CoV-2 variants within S protein besides VOC and COVID-19 outcomes, and 2) phylogenetic analysis of isolates from our region.

Methods: Whole genome sequences of SARS-CoV-2 was determined by Illumina MiSeq next-generation sequencer followed by phylogenetic analysis of full-genomes of SARS-CoV-2 from different regions.

Results: Clinical manifestations of 51 patients with COVID-19 were without any symptoms (13.7%), mild (47%), moderate (19.6%), severe (4%), critical (2%), and died (13.7%). Age of hospitalized patients (53.4 Â± 18 years) was higher than non-hospitalized patients (34.6 Â± 19) (p=0.001). A significant association between diabetes, hypertension, and anticoagulant and hospitalization was noted with p-value of 0.039 (OR=4.47 [95% CI=1.07-18.58]), 0.001 (OR=17 [95% CI=2-144]), and 0.02 (OR=27.97 [95% CI=1.54-507.13]), respectively; whereas a strong association between patientsâ□□ age, diabetes, anticoagulant, steroid and mortality was revealed with p-value of 0.016 (OR=8.44 [95% CI=1.5-47.49]), 0.019 (OR=8.5 [95% CI=1.43-50.66]), 0.001 (46.8 [95% CI=4.63-472.77]), and 0.009 (OR=15.75 [95% CI=2-123.86]), respectively. The viruses were classified as clade L (2%), GH (84.3%), GR (11.7%), and O (2%). Besides D614G, the most common variants in S protein were L5F (18.8%), V213A (18.8%), and S689R (8.3%). No significant association between multiple S protein variants and either hospitalization or mortality (p=0.11 and 0.69). Multivariate analysis showed that hypertension and anticoagulant were strong factors affecting hospitalization and mortality of COVID-19 patients with p-value of 0.009 (OR=17.06 [95% CI=2.02-144.36]) and 0.001 (OR=46.8 (95% CI=4.63-472.77), respectively. Interestingly, multiple S protein variants almost reached a significant level in affecting hospitalization (p=0.07). **Conclusions:** We showed for the first time the association between SARS-CoV-2 variants within S protein besides VOC and COVID-19 outcomes, revealing multiple S protein variants might affect COVID-19 severity, in addition to hypertension and anticoagulant. It further suggests the importance of genomic surveillance to monitor SARS-CoV-2 variants, particularly that might influence patientsâ□□ outcomes.

PrgmNr 2066 - Whole Exome Sequencing for Covid19 host in mild versus severe cases: A Saudi genome study

View session detail

Author Block: **H. Elbardis**[1,2], m. alaamery[3,4], S. Massadeh[3,4], Y. Arabi[5], M. Abedalthagafi[1,2], Saudi COVID-19 host genetics WES/WGS group; [1]King Abdulaziz City for Sci. and Technology (KACST), Riyadh, Saudi Arabia, [2]Genomics Res. Dept., Saudi Human Genome Project, King Fahad Med. City, Riyadh, Saudi Arabia, [3]Dev.al Med. Dept., King Abdullah Intl. Med. Res. Ctr. (KAIMRC), Riyadh, Saudi Arabia, [4]King Saud Bin Abdulaziz Univ. for Hlth.Sci., King Abdulaziz Med. City, Ministry of Natl. Guard Hlth.Affairs (MNGHA), Riyadh, Saudi Arabia, [5]Intensive Care Dept., Ministry of Natl. Guard Hlth.Affairs, King Saud Bin Abdulaziz Univ. for Hlth.Sci., King Abdullah Intl. Med. Res. Ctr.,, Riyadh, Saudi Arabia
Disclosure Block: **H. Elbardis:** None.
Background: The mortality and morbidity rates caused by the ongoing pandemic COVID-19 alerted the scientists to investigate the potential genetic patterns associated with symptom severity. From no or mild symptoms to severe life-threatening complications shed the light on the host genetic factors that incite the Covid-19 phenotypes. For Saudi Arabia, it was moderately affected by COVID-19, experiencing a case-fatality rate of 1.6 %, much lower than worldwide rate of 2.5 % according to WHO. In this perspective, exploring such genetic factors in Saudi population provides a deeper understanding of the heterogeneous phenotypes of the COVID-19 severity.Objective: We aimed to explore the genetic factors and variants associated with the reduced severity of symptoms seen in Saudi Arabian COVID-19 patients. Methods: Whole exome sequencing for 307 patients was performed where 149 patients were severe and 159 patients were mild. Exome library construction and Sequencing were done on ion torrent S5 XL system. Downstream analysis was performed to identify the common variants (MAF > 0.05), their allelic frequencies and gene ontology analysis in both severe and mild group. In silico validation of these variants was also performed.Results: Our study revealed 23 common genetic variants primarily connected to known co-morbidities which have not been reported to be of relevance to COVID-19 before. Gene ontology analysis revealed the involvement of a gene cluster associated with cytoskeletal functions. SNPs in 3 genes coding for keratin-associated proteins KRTAP10-4, KRTAP10-7, and Krtap5-5, are significantly associated with mild symptoms group, indicating a potential protective role.Conclusion: Our data highlights that the high level of autozygosity in this cohort not having apparent deleterious effects on mortality at the population level. The complete absence of common polymorphisms discovered previously to modulate COVID-19 symptomatology sheds the light on the genetic distinctiveness of different ethnicities. Our findings of the involvement of cytoskeletal components provides a better understanding of SARS-CoV pathogenesis, giving possible approaches for novel therapeutic targets.

PrgmNr 2067 - Advancing our understanding of genetic risk factors and potential personalized strategies in pelvic organ prolapse: currently largest GWAS reveals 19 novel associations

View session detail

Author Block: **N. Pujol Gualdo**[1], K. LÃ¤ll[1], M. Lepamets[1], H. Rossi[2], R. Arffman[2], T. Piltonen[2], R. MÃ¤gi[1], T. Laisk[1]; [1]Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [2]PEDEGO Res. Unit, Dept. of Obstetrics and Gynecology, Univ. of Oulu, Oulu, Finland
Disclosure Block: **N. Pujol Gualdo:** None.
Pelvic organ prolapse (POP) is characterized by a descent of the pelvic organs into the vaginal cavity. POP affects around 40% of women after menopause and is the most frequent reason for gynecological surgery. However, the etiology of POP remains to be elucidated and there is poor evidence guiding prevention and early detection of POP. In this study, we present the largest genome-wide association study (GWAS) meta-analysis using data from four European cohorts, including a total of 28,086 women with POP and 546,321 controls. We further integrate it with additional data layers, such as gene expression, to propose potential causal genes for the disorder. Moreover, for the first time, we aim to construct and validate the predictive ability of a polygenic risk score (PRS) alone or in combination with classical risk factors. Overall, we detected 26 genetic loci significantly associated with POP (pâ□□8), from which 19 loci were novel. The loci identified reinforces the role of connective tissue abnormalities (genes *EFEMP1*, *LOXL1*, *PLA2G6*), urogenital tract development (genes *WT1*, *DVL*, *WNT4*) and points towards an association with a range of cardiometabolic traits (genes *MAFF*, *LDAH*, *KLF13*), associations that were mirrored by genetic correlation analyses. The best fit PRS shows a hazard for incident POP of 1.63 times (95% CI: 1.46 to 1.79) higher in the top 5% PRS percentile compared to the remaining 95% percentile for the incident set, after adjusting for age. PRS in the incident subset shows higher predictive ability than clinical risk factors (Harrell C-statistic 0.583, sd= 0.007) and shows added value in combination with these (Harrell C-statistic 0.622, sd= 0.007), demonstrating that PRS in combination with clinical risk factors generates the best predictive model for POP. This study has the potential to improve our understanding of genetic factors underlying the polygenic architecture of POP and establishes the first evidence to base preventive strategies and early detection of POP including genetic risk factors.

PrgmNr 2068 - Explore the dynamic spatial and temporal regulation mechanism during embryonic sex development

View session detail

Author Block: **F. OU**[1], K. Miu[2], J. LIAO[1], H. SUEN[1], Z. WANG[1], W-Y. Chan[3]; [1]Sch. of BioMed. Sci., The Chinese Univ. of Hong Kong, Hong Kong, China, [2]The Chinese Univ. of Hong Kong, NT, China, [3]Flat D, 21st Floor, Ma On Shan, China
Disclosure Block: **F. Ou:** None.
Disorders of sex development (DSD) are congenital anomalies involving discordance between genotype to phenotype in gender identity. People with DSD suffer from ambiguous genitalia, impaired steroid hormone production, reduced or null fertility and partial to complete sex reversal, and many of them are vulnerable to severe psychological stress. Past genetic research mainly focused on proximal promoter activity or coding feature related to the diseases. Genes including *SRY*, *SOX9* and *NR5A1* have been reported to be critical for sex development. But the underlying genetic mechanism of most DSD is still unclear. With the advances in high throughput sequencing, stem cell and gene editing technologies, we aim to profile the spatial and temporal regulatory network based on scATAC-seq and complete stranded RNA-seq analysis and then validate our results in animal models and iPSC models. 3D chromatin topological interaction and non-coding RNA regulation are the research directions of interest, which may also help to explain the gene dosage regulation in development.

PrgmNr 2069 - Genetic characterization of the timing of human parturition unveils differential roles of maternal and fetal genomes on birth weight

View session detail

Author Block: **P. Sole-Navais**[1], C. Flatley[1], M. Vaudel[2], J. Chen[3], D. Westergaard[4], T. Laisk[5], Ã. Helgeland[2], R. N. Beaumont[6], M. E. Gabrielsen[7], M. Borges[8], A. Mahajan[9], G. G. Hayes[10], G-H. Moen[11], J. P. Bradfield[12,13], A. Skogholt[14], EGG Consortium, B. Feenstra[15], P. R. Njolstad[2], L. Muglia[16], R. Freathy[17], S. E. Johansson[2], G. Zhang[18], B. Jacobsson[1]; [1]Dept. of Obstetrics and Gynecology, Inst. of Clinical Sci., Sahlgrenska Academy, Univ. of Gothenburg, Gothenburg, Sweden, [2]Ctr. for Diabetes Res., Dept. of Clinical Sci., Univ. of Bergen, Bergen, Norway, [3]Cincinnati Childrens Hosp. Med. Ctr., Cincinnati, OH, [4]Novo Nordisk Fndn. Ctr. for Protein Res., Faculty of Hlth.and Med. Sci., Univ. of Copenhagen, Copenhagen, Denmark, [5]Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [6]Univ. of Exeter Med. Sch., Exeter, United Kingdom, [7]NTNU, Trondheim, Norway, [8]MRC Integrative Epidemiology Unit, Bristol Med. Sch., Univ. of Bristol, Bristol, United Kingdom, [9]Wellcome Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom, [10]Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [11]Oslo Univ. Hosp., Oslo, Oslo, Norway, [12]Quantinuum Res. LLC, San Diego, CA, [13]Ctr. for Applied Genomics, Childrenâ□□s Hosp. of Philadelphia, Philadelphia, PA, [14]NTNU, Trondheim, Trondheim, Norway, [15]Statens Serum Inst., Copenhagen, Denmark, [16]Cincinnati Children's Hosp. Med. Ctr., Cincinnait, OH, [17]Genetics of Complex Traits, The Coll. of Med. and Hlth., Univ. of Exeter, Exeter, United Kingdom, [18]Cincinnati Children S Hosp. Med. Ctr., Cincinnati, OH

Disclosure Block: **P. Sole-Navais:** None.

Introduction: The timing of human parturition is critical for neonatal survival and infant health, with preterm delivery (Methods: We performed a GWAS meta-analysis of gestational duration using maternal samples (n= 195K) and compared it to the genetic effects on birth weight and other female reproductive outcomes to disentangle the underlying genetics responsible for their relationship. The fetal effects on gestational duration and the maternal and fetal effects on birth weight were obtained from previously published GWAS meta-analysis. **Results:** We identified 24 independent loci (pADCY5 between the maternal effects on gestational duration and the fetal effects on birth weight. While the gestational duration index SNP (rs28654158) had a high population differentiation index (Africans and Europeans, Fst= 0.4, MAF= 0.01 vs 0.38, respectively), this was not the case for the birth weight index SNP (rs11708067); this is notable given the highest rate of preterm delivery in women from African ancestry. We observed significant genetic correlations with pre-eclampsia (r= -0.18, p= 2.1e-2) and testosterone in females (r= -0.24, p= 6.7e-5). **Conclusion:** Our results suggest that the maternal and fetal genomes shape fetal birth weight by acting on different phenotypes: timing of birth and growth rate, respectively.

PrgmNr 2070 - Genome-wide polygenic risk scores for hypertensive disease during pregnancy identify women at risk for long-term cardiovascular disease

View session detail

Author Block: **S. Lee**[1,2], M. Shivakumar[2], B. Xiao[2], S-H. Jung[2,3], Y. Nam[2], J-S. Yun[4,2], E. Choe[5], Y. Jung[1], C-W. Park[1], J. Park[1], J. Jun[1], D. Kim[2]; [1]Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Sungkyunkwan Univ., Seoul, Korea, Republic of, [4]The Catholic Univ. of Korea, Seoul, Korea, Republic of, [5]Seoul Natl. Univ. Hosp. Hlth.care System Gangnam Ctr., Seoul, Korea, Republic of

Disclosure Block: **S. Lee:** None.

Previous studies suggest that hypertensive disease during pregnancy (HDP) increases the risk of long-term cardiovascular disease later in life, and clinical guidelines recommend including HDP as an important female-specific factor in risk assessment. However, it has not been examined whether genetic predisposition for HDP affects the development of subsequent cardiovascular disease. In the current study, we developed polygenic risk scores for HDP (HDP-PRS) and evaluated its impact on long-term cardiovascular outcomes. We included unrelated European descent women (*n*=165,333) with at least one live birth and available genetic data from the UK Biobank. HDP-PRS was calculated using LDpred and summary statistics from the FinnGen Biobank. Subjects were divided according to the genetic risk categorized by HDP-PRS (High HDP-PRS (>75p) vs. low HDP-PRS (

PrgmNr 2071 - Large-scale GWAS meta-analysis unravels the genetic determinants of cervical biology and pathology

View session detail

Author Block: **T. Laisk**[1], M. Koel[1], U. VÃµsa[1], M. Lepamets[1], S. LemmelÃ¤[2], Estonian Biobank Research Team, FinnGen, H. M. Laivuori[3], M. J. Daly[2], P. Palta[1], R. MÃ¤gi[1]; [1]Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [2]Inst. for Molecular Med. Finland, Helsinki, Finland, [3]Univ Helsinki, Helsinki, Finland

Disclosure Block: **T. Laisk:** None.

The uterine cervix has an important role in female reproductive health, but not much is known about the genetic determinants of cervical biology and pathology. Here, we use data from large biobanks to characterise the genetics of cervical phenotypes (including cervical cancer) and leverage latest computational methods and gene expression data to refine the association signals for cervical cancer. Using Estonian Biobank and FinnGen data, we characterise the genetic signals associated with cervical ectropion (10,162 cases/151,347 controls), cervicitis (19,285/185,708) and cervical dysplasia (14,694/150,563). We present the results from the largest trans-ethnic GWAS meta-analysis of cervical cancer, including up to 9,229 cases and 490,304 controls from Estonian Biobank, the FinnGen study, the UK Biobank and Biobank Japan. We combine GWAS results with gene expression data and chromatin regulatory annotations in HeLa cervical carcinoma cells to propose the most likely candidate genes and causal variants for every locus associated with cervical cancer. We further dissect the HLA association with cervical pathology using imputed data on alleles and amino acid polymorphisms.

We report a single associated locus on 2q13 for both cervical ectropion (rs3748916, $p=5.1 \times 10^{-16}$) and cervicitis (rs1049137, $p=3.9 \times 10^{-10}$), and five signals for cervical dysplasia - 6p21.32 (rs1053726, $p=9.1 \times 10^{-9}$; rs36214159, 1.6×10^{-22}), 2q24.1 (rs12611652, $p=3.2 \times 10^{-9}$) near *DAPL1*, 2q13 ns1049137, $p=6.4 \times 10^{-9}$) near PAX8, and 5p15.33 (rs6866294, $p=2.1 \times 10^{-9}$), downstream of *CLPTM1L*. We identify five loci associated with cervical cancer: 1p36.12 (rs2268177, $p= 3.1 \times 10^{-8}$), 2q13 (rs4849177, $p=9.4 \times 10^{-15}$), 5p15.33 (rs27069, $p=1.3 \times 10^{-14}$), 17q12 (rs12603332, $p=1.2 \times 10^{-9}$), and 6p21.32 (rs35508382, $p=1.0 \times 10^{-39}$). Joint analysis of dysplasia and cancer datasets revealed an association on chromosome 19 (rs425787, $p=3.5 \times 10^{-8}$), near *CD70*.

Our results map *PAX8/PAX8-AS1*, *LINC00339*, *CDC42*, *CLPTM1L*, *HLA-DRB1*, *HLA-B,* and *GSDMB* as the most likely candidate genes for cervical cancer, which provides novel insight into cervical cancer pathogenesis and supports the role of genes involved in reproductive tract development, immune response and cellular proliferation/apoptosis. We further show that *PAX8/PAX8-AS1* has a central role in cervical biology and pathology, as it was associated with all analysed phenotypes. The detailed characterisation of association signals, together with mapping of causal variants and genes offers valuable leads for further functional studies.

PrgmNr 2072 - Meta-analysis of genome-wide association studies identified new loci associated with spontaneous preterm birth and gestational duration

View session detail

Author Block: **A. Pasanen**[1], M. K. Karjalainen[2], M. Ojaniemi[1], FinnGen, G. Zhang[3], L. J. Muglia[3,4], M. RÃ¤met[1], M. Hallman[1]; [1]PEDEGO Res. Unit, Med. Res. Ctr. Oulu, Univ. of Oulu, and Dept. of Children and Adolescents, Oulu Univ. Hosp., Oulu, Finland, [2]Ctr. for Life Course Hlth.Res., Faculty of Med., and Biocenter Oulu, Univ. of Oulu, Oulu, Finland, [3]Human Genetics Div., Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [4]Burroughs Wellcome Fund, Res. Triangle Park, Durham, NC
Disclosure Block: **A. Pasanen:** None.
Preterm birth (birth EBF1 and *EEFSEC*, with roles in B-cell development and selenium metabolism, were associated with both gestational length and SPTB. To complement the knowledge of the genetic background of SPTB, we conducted a fixed effects meta-analysis of gestational duration and SPTB in up to 98,372 women of European ancestry with data from FinnGen and two other cohorts. We detected several associated loci, of which eight were not previously associated with SPTB or the length of gestation. As an example, variants in *KCNAB1* intron were associated with gestational duration, and the association was replicated in 2,700 women from two Nordic birth cohorts. Noteworthy, *KCNAB1* variants were previously associated with offspring birth weight. Downstream in silico analysis suggested regulatory functions as potential causal mechanisms for the novel loci. The genome-wide association study in the FinnGen data replicated previous associations near *EBF1* and *EEFSEC*, among others.Altogether, our findings provide new insight into the genetic background of preterm birth. Better characterization of genetic pathways of SPTB has potential to suggest novel opportunities to treat and prevent preterm birth.

PrgmNr 2073 - Novel bi-allelic Mutations in *DNAH10* cause multiple morphological abnormalities of the flagella

View session detail

Author Block: **R. Zheng**; SiChuan Univ., Chengdu, China
Disclosure Block: **R. Zheng:** None.
Asthenozoospermia is characterized by attenuated sperm motility, further resulting in male infertility. Multiple morphological abnormalities of the sperm flagella (MMAF) is a typical asthenozoospermia, exhibiting various malformed flagellar shape. However, the known pathogenic genes can explain at most two-third of the MMAF cases. Herein, bi-allelic variants (c.9494C>G, [p.T3165R]; c.8378G>A, [p.R2793H]) of *DNAH10* were identified in a MMAF male by whole-exome sequencing (WES). Furthermore, by immunofluorescence staining and western blotting, we found these variants induced significant decrease of DNAH10 expression, further negatively impacting expression of DNAH2 in the spermatozoa from men harboring the *DNAH10* variant. Moreover, the outcomes of intracytoplasmic sperm injection (ICSI) were unsuccessful on the patient. Collectively, our data provided strong evidence to support that *DNAH10* is the causative gene of MMAF, and the finding of the novel mutations in *DNAH10* enriches the gene variants spectrum for MMAF which further contribute to diagnosis, genetic counseling and prognosis for male infertility.

PrgmNr 2074 - Disentangling migraine risk loci by fine-mapping and colocalization

View session detail

Author Block: **H. Hautakangas**[1], A. Palotie[1,2,3], M. Pirinen[1,4,5]; [1]Inst. for Molecular Med. Finland (FIMM), Helsinki Inst. of Life Sci. (HiLIFE), Univ. of Helsinki, Helsinki, Finland, [2]Analytic and Translational Genetics Unit, Dept. of Med., Dept. of Neurology and Dept. of Psychiatry Massachusetts Gen. Hosp., Boston, MA, [3]The Stanley Ctr. for Psychiatric Res. and Program in Med. and Population Genetics, The Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Dept. of Publ. Hlth., Univ. of Helsinki, Helsinki, Finland, [5]Dept. of Mathematics and Statistics, Univ. of Helsinki, Helsinki, Finland
Disclosure Block: **H. Hautakangas:** None.
Migraine is a highly prevalent brain disorder and genome-wide association studies (GWAS) have identified over hundred loci contributing to migraine risk, yet the causal variants and genes remain mostly unidentified. Here, we performed a Bayesian fine-mapping (FINEMAP) for the 123 risk loci identified in the latest migraine meta-analysis to narrow down potential causal variants in correlated association regions. The data sources included 23andMe, Inc. and UK biobank with up to 63,000 cases and 560,000 controls. Further, to prioritize candidate genes, we ran colocalization analyses by COLOC with expression quantitative trait loci (eQTL) datasets from multiple tissues from sources such as GTEx Consortium and eQTL Catalogue.

Examples of two loci where fine-mapping provided a clear picture are PHACTR1 locus on chromosome 6 and LRP1 locus on chromosome 12. In both loci, FINEMAP suggested only one causal variant with a posterior probability (PP) of around 70%. For PHACTR1 locus, the strongest candidate variant was rs9349379, an intron variant of *PHACTR1*, with PP = 1 and 95%-credible set (CS) containing only the named variant itself. The colocalization analyses also suggested very robust colocalization for *PHACTR1* in three arterial tissues (aorta, tibial and coronary arteries, all with PP = 1), while for other nearby genes colocalization was not supported (PP LRP1, forming the 95%-CS alone. Colocalization analyses suggested colocalization for *LRP1* in four tissues (aorta, tibial artery, heart ventricle and sun exposed skin, PP > 0.93).

Our goal is to extend these analyses to a larger migraine meta-analysis with over 100,000 cases, but the challenges are that we lack complete linkage disequilibrium (LD) information of the underlying GWAS data and that there is variation in effective sample sizes (N_e) across variants. Therefore, we have evaluated how fine-mapping with an LD reference panel (from UK Biobank) performs. We observed that it was necessary to restrict the analysis to variants with similar N_e, but after that, the results using the reference panel were comparable to the results using in-sample LD.

To conclude, these results help to prioritize causal variants and genes in migraine risk loci for future translational research on migraine.

PrgmNr 2076 - Long-read sequencing analysis of Alzheimer risk gene*ABCA7* identifies risk-increasing novel alternative splicing events

View session detail

Author Block: **L. Duchateau**[1], F. KÃ¼Ã§Ã¼kali[1], J. Van Dongen[1], L. Bossaerts[2], E. Hens[2,3], A. De Roeck[1], C. Van Broeckhoven[2,3], K. Sleegers[1]; [1]Complex Genetics Of Alzheimer's Disease group, VIB-UAntwerp Ctr. for Molecular Neurology, Wilrijk, Belgium, [2]Neurodegenerative Brain Diseases group, VIB-UAntwerp Ctr. for Molecular Neurology, Wilrijk, Belgium, [3]Dept. of Neurology, Univ. Hosp. Antwerp, Antwerp, Belgium

Disclosure Block: **L. Duchateau:** None.

ABCA7 (*ATP-Binding Cassette Subfamily A member 7*) is an Alzheimerâ□□s disease (AD) risk gene and encodes for a lipid transporter. Studies revealing novel splice events, including â□□rescue splicingâ□□ events that are able to rescue the effect of premature termination codon (PTC) mutations, suggest that the *ABCA7* splicing profile is complex. These splicing differences could potentially account for the large differences found in *ABCA7* expression inter-individually as well as differences in phenotype and age at onset (AAO) in mutation carriers. To study the *ABCA7* transcript, we used targeted longâ□□read cDNA sequencing on an Oxford Nanopore MinION platform in a unique population of 39 *ABCA7* PTC carriers and 10 non-mutation carriers, including 28 AD patients, 17 controls and 5 mild cognitive impairment patients. RNA was extracted from 47 lymphoblastoid cell lines, 4 hippocampi and 2 prefrontal cortices. Using the FLAIR (full-length alternative isoform analysis of RNA) pipeline and an in-house R script, alternative splicing was quantified and studied. In total we identified 199 previously undescribed splice events with an alternative/canonical ratio of at least 0.05, an almost 4-fold increase compared to the 52 known junctions. We found that both AD patients and PTC mutation carriers had a significantly higher degree of reads containing alternative splice events compared to controls (p=2.36*10-11 and pABCA7 splicing is much more complex than initially thought and is significantly increased in AD patients and outside of regions of structural or functional importance. This suggest that alternative splicing may affect ABCA7 function and through this have an influence on AD risk. Our data suggests that having a PTC mutation in *ABCA7* might increase susceptibility to (rescue) splicing.

PrgmNr 2077 - Polygenic risk scores as a marker for lifetime epilepsy risk

View session detail

Author Block: **H. O. Heyne**[1,2], FinnGen, M. J. Daly[3,1], R. KÃ¤lviÃ¤inen[4]; [1]Inst. for Molecular Med. Finnland: FIMM, Helsinki, Finland, [2]Hasso Plattner Inst., Potsdam, Germany, [3]Massachusetts Gen. Hosp., Boston, MA, [4]Kuopio Univ. Hosp., Kuopio, Finland
Disclosure Block: **H.O. Heyne:** None.
Epilepsy affects approximately 1% of individuals worldwide. Making an epilepsy diagnosis is often difficult with estimates that up to 25% of epilepsy could initially be misdiagnosed. The SNP heritability of genetic generalized epilepsy is high (32%) and it has recently been shown that individuals with epilepsy also have elevated epilepsy polygenic risk scores (PRS). However, investigation how epilepsy PRS may predict epilepsy risk in an individual across lifetime or in individuals with unclear seizure events has so far been lacking.Here, we studied epilepsy PRS in detailed longitudinal electronic health records of > 269k Finns including ICD codes and drug purchases of over 50 years. Our dataset included 9660 individuals with epilepsy related diagnoses. We could confirm previous studies describing an elevated PRS for generalized epilepsy (PRS_{gen}) in individuals with generalized epilepsy. This was particularly high for juvenile myoclonic epilepsy, which could be because it represented the largest diagnosis group of the GWAS that was used to construct the PRS. Individuals with a top 10% PRS_{gen} in FinnGen had a more than doubled lifetime risk (p-value $1x10^{-11}$, hazard ratio 2.3) to develop generalized epilepsy compared to the bottom 90% PRS_{gen}. We further found that over half of individuals with specific diagnoses of generalized or focal epilepsy were initially diagnosed with unclear convulsions (R56.8) or unclear epilepsy (G40.9). Their PRS_{gen} was significantly higher than of those individuals who had only one unclear seizure event and did not later receive an epilepsy diagnosis. Specifically, individuals with a top 10% PRS_{gen} had a hazard ratio of 2.4 (p-value $1x10^{-4}$) to progress to generalized epilepsy after an unclear seizure event below the age of 40 compared to the bottom 90% PRS_{gen}. These results indicate a future potential for epilepsy PRS to help in predicting progression to epilepsy.

PrgmNr 2078 - Protein interaction network reveals enriched genetic variation across signaling networks in frontotemporal dementia

View session detail

Author Block: **C. Kocoglu**[1,2], R. Ferrari[3], M. Roes[1,2], G. Vandeweyer[4], F. Kooy[4], C. Van Broeckhoven[1,2], C. Manzoni[5], J. van der Zee[1,2]; [1]Neurodegenerative Brain Diseases, VIB Ctr. for Molecular Neurology, VIB, Antwerp, Belgium, [2]Dept. of BioMed. Sci., Univ. of Antwerp, Antwerp, Belgium, [3]Dept. of Neurodegenerative Disease, UCL, London, United Kingdom, [4]Dept. of Med. Genetics, Univ. of Antwerp, Antwerp, Belgium, [5]Sch. of Pharmacy, UCL, London, United Kingdom
Disclosure Block: **C. Kocoglu:** None.
Frontotemporal dementia (FTD) is a complex early-onset neurodegenerative brain disorder. Several disease-causing genes have been identified for distinct clinical and pathological subtypes of familial FTD mainly via linkage analysis in large pedigrees. Yet, the genetic etiology of sporadic FTD is poorly understood and complicated by the highly heterogeneous disease presentation. To further unravel the complex genetic etiology on non-Mendelian FTD, we hypothesized that interactors of the protein products of known FTD genes might also play a role in disease onset. We therefore applied protein network analysis updating and automatizing an existing weighted protein-protein interaction network analysis (WPPINA) pipeline (Ferrari et al., 2017) to prioritize candidate genes for downstream rare variant association analysis on exome data. We created an FTD protein interaction network (**FTD-PIN**) starting from 14 known Mendelian FTD genes. Using these FTD genes as seeds, we downloaded 1st- and 2nd-degree physical interactors via the PINOT online server. We performed enrichment analysis for Gene Ontology - biological processes (GO:BP) terms on the genes from the FTD-PIN using g:Profiler. Enriched GO:BP terms were grouped into custom semantic classes and these were further prioritized based on weighted number of GO-BPs as a function of the term *P*-value and term size. This analysis identified the following most **overrepresented FTD processes**: *waste disposal, response to stimulus, immune system* and *cell death*. The FTD-PIN genes belonging to these overrepresented FTD processes made up the **prioritized gene list** (n=442). On a Belgian WES dataset of 228 FTD patients and 345 controls, we multiple rare variant association tests with the rvtest package on the prioritized gene list. Four genes showed suggestive enrichment of missense variants in all statistical models: *TNFAIP3*, *DNM2*, *RARA* and *UBR4* (*P* TNFAIP3 as the top gene with $P = 0.0007$, reaching near test-wide significance ($p=2.5x10^{-4}$). We subsequently extracted the TNFAIP3-interactome from the FTD-PIN and performed enrichment of GO:BP terms on this subset (n=50) to identify possible disease processes of TNFAIP3 which pointed towards signalling processes, specifically: *cell death - signalling, immune system - signalling - toll-like, immune system - signalling - cytokine*. Our study indicates that integration of protein interaction network and functional enrichment analyses is a useful approach to increase power to identify rare variants with low-effect size and genes contributing to the disease biology of FTD.

PrgmNr 2079 - Trans-ethnic fine-mapping in the major histocompatibility complex region on Parkinson's disease risk

View session detail

Author Block: **T. Naito**[1,2], W. Satake[2], K. Ogawa[1], K. Suzuki[3], J. Hirata[4], J. N. Foo[5], E. K. Tan[6], T. Toda[2], Y. Okada[7]; [1]Osaka Univ. Graduate Sch. of Med., Suita, Japan, [2]The Univ. of Tokyo, Tokyo, Japan, [3]Osaka Univ. Graduate Sch. of Med., Suita, Japan, Suita, Japan, [4]Teijin Pharma Limited, Tokyo, Japan, [5]Genome Inst Singapore, Singapore, Singapore, [6]Natl. NeuroSci. Inst., Singapore Gen. Hosp., Singapore, Singapore, [7]Osaka Univ. Graduate Sch. of Med., Suita, Osaka, Japan
Disclosure Block: **T. Naito:** None.
Although different studies reported the genetic predisposition of human leukocyte antigen (HLA) to the risk of Parkinsonâ□□s disease (PD), the complex haplotype structure and highly polymorphic feature of the major histocompatibility complex (MHC) region has hampered a unified insight on the causal HLA variants. In addition, most previous studies focused on European populations and the evidence for non-European populations has been scarce. We conducted trans-ethnic fine-mapping for large cohorts to elucidate shared and distinct genetic features of the MHC region associated with PD risk. We targeted European populations (14,650 cases and 1,288,625 controls) and East Asian populations (7,712 cases and 27,372 controls). We adopted a hybrid fine-mapping approach including: (1) HLA genotype imputation of GWAS genotype data using our novel imputation method based on deep learning, DEEP*HLA; and (2) direct imputation of HLA variant risk from the GWAS summary statistics. The strongest association was observed at the protective effect of His13 in HLA-DRÎ²1 (P = 6.0 Ã□ $10^{â□□15}$), and this position explained the majority of the risk in *HLA-DRB1*. *In silico* prediction revealed that *HLA-DRB1* alleles with His13 (protective) and Arg13 (risk) had significantly weaker and stronger binding-affinity to an Î±-synuclein epitope than other alleles (P = 9.6 Ã□ $10^{â□□4}$ and 1.0 Ã□ $10^{â□□3}$, respectively). It might suggest that HLA-DRÎ²1 position 13 associates with the etiology of PD by altering the binding-affinity to the Î±-synuclein epitope and determining subsequent immune responses. Stepwise conditional analysis revealed additional independent association at Ala69 in HLA-B (P = 1.0 Ã□ $10^{â□□7}$). ï»¿Furthermore, a sub-analysis in Europeans suggested additional independent associations at non-HLA genes in the class III MHC region (*EHMT2*; P = 2.5 Ã□ $10^{â□□7}$) The current study highlights the genetic features of the MHC region associated with the risk of PD shared across ethnicities, enhancing our understanding of the immunologic pathophysiology of PD.

PrgmNr 2080 - Using non-Autologous Wharton's Jelly Mesenchymal Stem Cells and Chitosan/Poly ethylene oxide as a Synthetic Scaffold inhibit SNI-Induced Apoptosis in Rat

View session detail

Author Block: **M. Moattari**[1], F. Moattari[2]; [1]KhU, Tehran, Iran, Islamic Republic of, [2]Persian Gulf Univ., Bushehr, Iran, Islamic Republic of
Disclosure Block: **M. Moattari:** None.
Introduction: Inflammation, apoptosis and oxidative stress are considered to be important physiological conditions associated with the development of sciatic nerve injury. We hereby add our experience in Neuroscience research center of a surgical reconstruction of the damaged sciatic nerve using a mixture of SCs and Chitosan/Poly ethylene Oxide (Cs/PEO) as a synthetic scaffold by using TUNEL assay. Materials and Methods: Thirty rats of severely damaged sciatic nerve have been operated for end-to-end suture and reconstruction of the sciatic nerve using SCs and Chitosan/poly ethylene oxide (Cs/PEO). Cs/PEO serves to maintain the position of the SCs. An average of 5 Ã 104 of non-autologous Wharon's jelly SCs. The in-situ DNA fragmentation was visualized by terminal deoxynucleotidyl transferase dUTP nick (TUNEL) assay. Results: Our results showed that 8 weeks after surgery, sciatic nerve injury significantly increased the percentage of apoptotic cells which was inhibited by a SCs and Chitosan/poly ethylene oxide (Cs/PEO) as reflected by TUNEL results (p

PrgmNr 2082 - Assessing association between polygenic risk scores for mood disorders and substance involvement in the Taiwan Biobank

View session detail

Author Block: **R-Y. Lai**[1], S-H. Wang[2,1]; [1]China Med. Univ., Taichung, Taiwan, [2]Taichung, Taiwan
Disclosure Block: **R. Lai:** None.
Background: The comorbidity of mood disorders, including bipolar disorder (BPD) and major depressive disorder (MDD), with substance use was common. In addition to a bidirectional relationship proposed by prospective studies and Mendelian randomization studies, a shared pathophysiology, such as common genetic architecture, could also contributed to the comorbidity. This study aimed to explore the association of polygenic liabilities for BPD and MDD with a range of substance use phenotypes by conducting a population-based study in Taiwan.
Method: The study subjects were recruited from the Taiwan Biobank. Genome-wide genotyping data was available for 77088 unrelated community participants without mood disorders. We applied polygenic architecture profiling and calculated polygenic risk score (PRS) for BPD and MDD in each individual. Data for substance involvement included tobacco, alcohol, and betel nut use phenotypes. Detailed information in terms of age at initiation, duration and current status was collected for participants who had used a substance. A regression model with adjustments for sex, age, and population stratification components was performed to estimate the strength of association of PRS with substance involvement. Sex difference in the PRS association was also explored.
Results: The MDD PRS was positively associated with regular alcohol use (OR in per SD increase in PRS=1.03, p=0.02), ever tobacco use (OR=1.05, p**Conclusion:** This study supported that the shared genetic architecture contributed to the comorbidity between MDD and substance involvement. Our results also suggested sex difference in genetic prediction, women with MDD susceptibility had a higher risk of tobacco and betel nut involvement than men.

PrgmNr 2083 - Brain eQTL of East Asian, African American, and European Descent Explains Schizophrenia GWAS in Diverse Populations

View session detail

Author Block: **Y. Chen**[1], S. Liu[2], F. Wang[1], Y. Jiang[3], Y. Xia[4], W. Qiu[5], C. Ma[5], X. Yan[1], J. Huang[1], S. Xu[6], B. Tang[1], H. Huang[7], C. LIU[8], C. Chen[9]; [1]Central South Univ., changsha, China, [2]Central South Univ., China, China, [3]Nashville, [4]Broad Inst., Medford, MA, [5]Chinese Academy of Med. Sci. and Peking Union Med. Coll., Beijing, China, [6]CAS-MPG Partner Inst. for Computational Biology, Shanghai, China, [7]Boston, MA, [8]SUNY Upstate Med. Univ., Syracuse, NY, [9]Central South Univ., Changsha, China
Disclosure Block: **Y. Chen:** None.
Previous studies integrating genome-wide association studies (GWAS) and brain expression quantitative trait loci (eQTL) data have discovered hundreds of risk genes associated with schizophrenia (SCZ). However, most studies focused on individuals of European Populations. Based on the differences in population structures such as linkage disequilibrium patterns and allele frequencies, we hypothesized that different populations had their own SNPs associated with brain gene expression; and population-relevant brain eQTL will improve interpretation of GWAS signals of the corresponding populations than the use of eQTL from other populations in SCZ.We first characterized the eQTL using RNA-seq and WGS data of brain samples from subjects of East Asian (EA, N = 228), Africa American (AFR, N = 175), and European (EU, N = 407) ancestries. Using FDR q-value Population shared (PS-), defined as the eQTL, of the identical SNP-gene pair, with the same effect in all three populations; or *Population-unique* (PU-), defined as the eQTL is only significant or has significantly different effect size in one population. Striking differences existed in eQTL across populations: only 3,999 genes were significantly associated with the same SNPs in three populations. We identified 1,308 and 333 PU-eQTL Genes in EA and AFR, respectively. Meanwhile, the SNP minor allele frequency and fixation index values of eQTLs are significantly higher in the PU-eQTL SNP than the PS-eQTL SNP (adjusted p

PrgmNr 2084 - Contribution of Glucoside Xylosyltransferase 1 gene variants in Schizophrenia

View session detail

Author Block: **B. Shekhar**; ICMR-Natl. Inst. for research in reproductive Hlth., Mumbai, India
Disclosure Block: **B. Shekhar:** None.
Schizophrenia (SCZ) is a chronic neuropsychiatry disorder affecting true perception of an individual. Impaired social interactions along with Hallucination and Delusions are the hallmark of the disorder. Polygenic and heterogenous nature of the disorder make it challenging to confer any concrete evidence for the onset of SCZ. Although, variant based genome wide association studies and rare variants detection using Next Generation Sequencing have helped to decipher numerous variants that are responsible for the onset of SCZ, but still various links are missing in order to understand this neuropsychiatry disorder in a better way. Using Whole Exome Sequencing we identified 4 heterozygous variants in Glucoside Xylosyltransferase 1 gene (GXYLT1, 12q12) having 8 exons in affected members of 6 families. These variants are predicted as damaging according to various prediction tool. Upon detailed analysis these 4 variants viz: rs79044728, rs1262821887, rs1462576913 and rs796100580 are found to be altering proteins as p.Y265C, p.G33X, p.Y264N and p.E249G. Out of these 4 protein variants one variants prematurely truncate the protein. Glucoside Xylosyltransferase 1 is a member of Glycosyl Transferase 8 family. GXYLT1 adds first xylose to O-glucose residues in 6 cysteines repeat of epidermal growth factor-like (EGF) repeats. EGF repeats are present in the extracellular domain of NOTCH and are conserved throughout the species. NOTCH signaling pathway has prominent role in neurodevelopment and adult brain homeostasis. It assists neurodevelopment and patterning by regulation of neurogenesis, axonal growth and synaptogenesis. There is substantial number of evidences present suggesting the role of NOTCH signaling pathway in the onset of SCZ. Also, it has been shown that the addition of Xylose to O-glucose has inhibitory effect on NOTCH signaling pathway. This study delineates the role of GXYLT1 in the onset of Schizophrenia through NOTCH signaling pathway that regulate neural patterning and neurogenesis in adult brain.

PrgmNr 2086 - Is chronotype a risk factor for neuropsychiatric disorders? A two-sample, multivariable Mendelian randomisation study

View session detail

Author Block: **S. Crinion**[1], L. Lopez[2], D. Morris[1]; [1]NUI Galway, Galway, Ireland, [2]Maynooth Univ., Maynooth, Ireland
Disclosure Block: **S. Crinion:** None.
Disruption of circadian rhythm is a common feature in many neuropsychiatric disorders including autism and schizophrenia. Chronotype, an individual's synchronisation to the 24 hour day, is commonly used as a proxy for circadian rhythm disruption. Being a morning person, someone who prefers waking and going to bed earlier, is genetically correlated with increased well-being and decreased risk of neuropsychiatric disorders. We performed a two-sample Mendelian randomisation (MR) study to determine the effect of chronotype on risk of six neuropsychiatric disorders (autism spectrum disorder, attention deficit hyperactivity disorder, bipolar disorder, insomnia, major depressive disorder and schizophrenia). We used 351 independent genome-wide significance loci (p

PrgmNr 2087 - Major depressive disorder and current psychological symptoms modify the polygenic predisposition to body mass index on obesity-related traits

View session detail

Author Block: **S-H. Wang**, C-Y. Su; China Med. Univ., Taichung, Taiwan
Disclosure Block: **S. Wang:** None.
Background: Major depressive disorder (MDD) and obesity are both leading serious public health problems and highly prevalent in modern society. A bidirectional relationship between MDD and obesity have been evidenced. Their comorbidity or complications might be attributable to genetic influence. Specific genetic variants for body mass index (BMI) have been shown to exhibit larger effect sizes in depressed individuals. This study aimed to explore the association of polygenic predisposition with obesity-related traits, and to examine whether such polygenic influence modified by MDD. **Methods:** The study subjects were recruited from Taiwan Biobank. Genome-wide genotyping was available in 80307 unrelated individuals. We applied polygenic architecture profiling and calculated the polygenic risk score (PRS), which refers to the cumulative additive effect of trait-associated variants across the genome, for BMI and for MDD. The PRS was then normalized to a Z score. Obesity-related outcomes included BMI, overweight (BMI â¥25), obesity (â¥30), waistline, hipline, waist-hip ratio (WHR) and body fat rat. MDD was measured by self-reported lifetime diagnosis and current psychological symptoms was measured by Patient Health Questionnaire-4 (PHQ-4). The significance of the PRS influence was evaluated using a linear regression model or a logistic regression model with adjustments for gender, age and population stratification components. The modified effects of MDD and current psychological symptoms were evaluated by an interaction term. The interaction between PRS for MDD and PRS for BMI was also examined. **Results:** The MDD PRS was positively associated with waistline (beta in per SD increase in PRS=0.12), hipline (beta=0.09), WHR (beta=0.04), body fat rate (beta=0.09), BMI (beta=0.05), overweight (OR=1.03), and obesity (OR=1.06). There was a significant interaction between BMI PRS and MDD and between BMI PRS and current psychological symptoms on obesity-related traits. The effect of BMI PRS on body fat rate was larger for individuals with MDD (beta=1.19 v,s. beta=0.98 for those without). The effect of BMI PRS on BMI was larger for individuals with MDD (beta=0.83 v,s. beta=0.70 for those without). The amplified effect of BMI PRS was also observed for individuals with current psychological symptoms. There was a synergistic interaction between MDD PRS and BMI PRS on obesity. **Conclusions:** Higher polygenic predisposition to MDD was associated with obesity. Lifetime MDD and current psychological symptoms amplify the effect of polygenic predisposition to BMI on obesity.

PrgmNr 2088 - A phenome-wide association study of Y-chromosomal haplogroups and 28 disease endpoints in 37,518 Finnish men

View session detail

Author Block: **A. Preussner**[1], J. T. Leinonen[1], S. Kerminen[1], M. Pirinen[1,2,3], FinnGen, T. Tukiainen[1]; [1]Inst. for Molecular Med. Finland (FIMM), Univ. of Helsinki, Helsinki, Finland, [2]Dept. of Publ. Hlth., Univ. of Helsinki, Helsinki, Finland, [3]Helsinki Inst. for Information Technology HIIT and Dept. of Mathematics and Statistics, Univ. of Helsinki, Helsinki, Finland
Disclosure Block: **A. Preussner:** None.
The Y chromosome is routinely excluded from genetic association studies due to the peculiar biology and analytical challenges specific to the Y chromosome. Consequently, potential impacts of Y-chromosomal genetic variation on complex disease remains largely uncharacterized.
To explore the role of the Y chromosome in complex disease, we performed a phenome-wide association study (PheWAS) of Y-chromosomal haplogroups in 37,518 Finnish men (24,160 for discovery + 13,358 for replication) from the FinnGen project. The analysis focused on the four most common haplogroups in Finland (N1c1, I1, R1a, R1b) and 28 common disease endpoints (â¥3000 cases in the dataset) selected from ICD-hierarchy level 2, e.g. hernias, hypertension and dorsopathies. As the haplogroup frequencies, for N1c1 in particular, were correlated with the autosomal genome, the logistic regression analyses were carefully adjusted for population structure in addition to other covariates.
Seven nominally significant associations were observed in the discovery analysis, and three of these associations remained nominally significant in the meta-analysis. All three replicated associations included cardiometabolic system diagnoses, namely both risk increasing and protective effects for ischemic heart disease (OR = 0.93, p = 0.023 for N1c1; OR = 1.08, p = 0.023 for I1), and a protective effect for diabetes (OR = 0.87, p = 0.018 for R1a). However, none of the associations remained statistically significant after correcting for multiple testing (p Nevertheless, the finding that haplogroup I1 (frequency 26% in Finland) may increase the risk for ischemic heart disease in Finland supports the previously reported link between haplogroup I1 and increased risk for coronary artery disease (OR = 1.11 in the UK), suggesting the Y chromosome contains robust disease associations. The potential protective role of haplogroup N1c1 (frequency 60% in Finland) further suggests the Y chromosome may harbor additional genetic variations that impact heart disease susceptibility. Overall, with this pilot study, which to our knowledge is the largest PheWAS of the Y chromosome including Finnish haplogroups, we show the feasibility of including Y-chromosomal data in genetic association analyses. The few suggestive associations observed propose further investigations and validation in larger datasets are warranted to define the speculated role of Y chromosome in complex disease.

PrgmNr 2089 - A selection pressure landscape for 870 human polygenic traits

View session detail

Author Block: **G. Lin**, W. Song, W. Wang; Shanghai Jiao Tong Univ., Shanghai, China
Disclosure Block: **G. Lin:** None.
Characterizing the natural selection of complex traits is essential for understanding human evolution, and both biological and pathological mechanisms. To fulfill this requirement, we leveraged genome-wide summary statistics for 870 polygenic traits and attempted to comprehensively quantify the selection pressure on traits of different forms in European ancestry across four different human development times. We found that 88% of traits underwent polygenic adaptation in the past 2000~3000 years. The selection at a later time was generally dependent on ancient selection pressures of the same trait. Traits related to pigmentation, body measurement, and nutritional intake exhibited strong selection signals across different time scales. In sum, we provided a first-step global overview of natural selection on human polygenic traits and their essential characteristics across human evolution, which could serve as a foundation for further populational and medical genetic studies.

PrgmNr 2090 - Analysis of genetic variants of SARS-CoV-2 host factors *ACE2*, *NRP1*, *TMPRSS2*, and *FURIN* in Japanese population from IRUD whole-exome sequencing data

View session detail

Author Block: **K. Satou**[1], K. Yanagi[1], M. Omata[1], A. Igarashi[1], T. Hidai[1], Y. Matsubara[2], T. Kaname[1]; [1]Dept. of Genome Med., Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan, [2]Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan
Disclosure Block: **K. Satou:** None.
Background: Coronavirus disease 2019 (COVID-19) is an ongoing global pandemic infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first case was identified in China in December 2019 and a cumulative total of more than 176 million cases and 3.8 million deaths have been reported worldwide to date. Patients with the COVID-19 present a broad spectrum of clinical presentation and a large number of host risk factors are reported, including age, sex, underlying condition, and local population. The SARS-CoV-2 utilizes host (human) factors, including angiotensin converting enzyme 2 (*ACE2*), neuropilin 1 (*NRP1*), transmembrane serine protease 2 (*TMPRSS2*), and paired basic amino acid cleaving enzyme (*FURIN*), in its cell entry stage. Revealing genetic variant profiles among individuals or populations is essential for developing precision medicine.
Methods: In this study, we analyzed genetic variants of the SARS-CoV-2 host factor genes by using whole-exome sequencing data of 2,048 healthy individuals in Japanese population. We utilized the data from the Initiative on Rare and Undiagnosed Diseases (IRUD) project, which we are currently working on. Allele frequencies of the variants in the Japanese population were compared to local populations in the public database. Pathogenicity of the variants were evaluated by prediction score algorithms. Exonic nonsynonymous variants existing on virus-host protein-protein interaction surface were also evaluated for selected proteins by using molecular modeling.
Result: We found 66 variants for *ACE2* and 9 were exonic and nonsynonymous, 220 variants for *NRP1* and 15 were exonic nonsynonymous and 1 were stopgain, 223 variants for *TMPRSS2* and 19 were exonic nonsynonymous, and 112 variants for *FURIN* and 22 were exonic nonsynonymous. More than half of the variants found in each gene were not available in the database. The allele frequency of some variants showed differences among local populations.
Conclusion: We analyzed genetic variants of SARS-CoV-2 host factors *ACE2*, *NRP1*, *TMPRSS2*, and *FURIN* in Japanese population from IRUD whole-exome sequencing data. We found total 621 variants for four genes. More than half of the variants were not available in the public database and the allele frequencies were not sufficiently comparable. Revealing genetic variant profiles among individuals or populations is essential for developing precision medicine. We hope that our result will contribute as one of the basic data for developing COVID-19 genome medicine.

PrgmNr 2091 - Analysis of the genetic components behind the chronic back pain

View session detail

Author Block: **E. Elgaeva**[1,2], M. Freidin[3], F. Williams[3], P. Suri[4,5], Y. S. Aulchenko[6,1], Y. Tsepilov[1,2]; [1]Inst. of Cytology and Genetics, Siberian Branch of Russian Academy of Sci., Novosibirsk, Russian Federation, [2]Novosibirsk State Univ., Novosibirsk, Russian Federation, [3]Dept. of Twin Res. and Genetic Epidemiology, Sch. of Life Course Sci., King's Coll. London, London, United Kingdom, [4]Univ. of Washington, Seattle, WA, [5]VA Puget Sound Hlth.Care System, Seattle, WA, [6]PolyOmica, Maastricht, Netherlands

Disclosure Block: **E. Elgaeva:** None.

Background: Back pain is a highly prevalent disabling condition with a significant socioeconomic burden. In 10% of cases acute back pain converts to a chronic condition. Polygenicity and heritability up to 68% have been shown for this trait. Furthermore, there is evidence for shared genetic background of various pain conditions. Recently, ~80% of chronic musculoskeletal pain heritability was found to be shared across four locations. It is thought that shared genetic background may involve the psychosocial component of chronic pain as a major component. Similarly, an existence of genetic factors that are unique to chronic pain at specific locations can be assumed. Here we used novel methods to study the genetic factors shared by chronic pain traits and the unique genetic factors of chronic back pain (CBP). **Materials and methods**: The current study was performed using the GWAS results on various chronic pain traits (back, neck, hip, knee, stomach pain and headache). All the data were obtained from UK Biobank project #18219 (N = 456,000) and split into a discovery (N = 265,000, Europeans) and a replication set (N = 191,000, Africans, South Asians and Europeans). Firstly, for all traits we assessed phenotypic correlations as described by Stephens et al. and calculated genetic correlations and heritability using LD Score regression. We used these parameters to build the optimal linear combination of traits, maximizing their shared heritability (MaxSH methodology) and obtained GWAS results for the shared genetic component (SGC) of chronic pain traits. Finally, we adjusted CBP using SGC and obtained new GWAS data for the unique genetic component (UGC) of CBP. Loci associated with SGC or UGC of CBP at p Results: We isolated SGC of chronic pain traits and UGC of CBP. Nine loci were shown to be associated with SGC, five of them were observed earlier in Tsepilov et al. (2020) study and four loci were new with two of them replicated. Similarly, we identified one locus significantly associated with UGC of CBP. This association was previously found in a back pain GWAS by Freidin et al. in 2018. As expected, the regions associated with SGC were enriched for genes expressed in the nervous system, while regions associated with UGC of CBP were related to the musculoskeletal system. Hence, SGC is likely to primary reflect psychosocial aspects of CBP and chronic pain traits in general, while UGC primarily reflects the anatomical basis of CBP.

PrgmNr 2092 - Association of confounding factors with Adolescent Idiopathic Scoliosis in the population of Jammu andKashmir, India

View session detail

Author Block: **H. Singh**[1], S. Kowra[1], N. Gupta[2], G. Gupta[3], S. Pandita[4], V. Singh[1], E. Rai[5], S. Ikegawa[6], S. Sharma[7]; [1]Sch. of Biotechnology, Shri Mata Vaishno Devi Univ., Katra, India, [2]Dept. of Hlth.and Family Welfare, Jammu, India, [3]Dept. of Radiology, Shri Mata Vaishno Devi Narayana Superspeciality Hosp., Katra, Jammu, India, [4]Radiology section, Chest Disease Hosp., Bakshi Nagar, Jammu, India, [5]Shri Mata Vaishno Devi Univ., Katra, India, [6]IMS, RIKEN, Tokyo, Japan, [7]SHRI MATA VAISHNO DEVI Univ., KATRA, India

Disclosure Block: **H. Singh:** None.

Adolescent idiopathic scoliosis (AIS) is a complex 3D spinal deformity characterized by Cobb angle â¥ 10Â°. It affects 1-3% of adolescent population globally. In our study, 9500 adolescent individuals (5,001 males and 4,499 females) were screened for scoliosis in the population of Jammu region of Jammu and Kashmir. Anthropometric measurements including height, weight and BMI were recorded. Screening of the individuals were performed using Scoliosis-meter was that provides the measure of angle of trunk rotation (ATR), where ATR of 7Â° were suspected to be scoliotic and further examined radiologically. The prevalence of AIS in our population was observed to be 0.60% which is less than the global average estimate but greater than the previously reported studies in India. We tried to statistically evaluate the correlation and association of anthropometric factors with AIS in our study. The anthropometric factors were found to be uncorrelated correlated with ATR (r= -0.071, p value= 0.580, r= -0.104, p value=0.415 and r = -0.129, p value= 0.314 for BMI, height and weight respectively). Further, we have also tried to evaluated the association of anthropometric factors with AIS using independent t-test for BMI, height and weight among the AIS patients and healthy controls. No significant association was observed between anthropometric factors and AIS in overall screened population. However, BMI showed significant association with p value = 0.031 among AIS patients and the healthy controls in the age group of 12-16 years whereas no significant association has been observed for other anthropometric factors including height and weight in this age group with p-value of 0.251 and 0.304 respectively. The level of significance between BMI and AIS is not very high so, we can confer that BMI may act as a confounding factor for AIS but is not only a single contributory factor for the AIS predisposition. It seems that genetic factors might be a major contributor in the pathogenesis of the AIS along with some confounding factors such as BMI, height and weight. Therefore, it becomes pertinent to carry out genetic studies on AIS in our population group to get a clear picture of AIS predisposition.

PrgmNr 2093 - Can sex differences in GWAS variant effects be found with current sample sizes?

View session detail

Author Block: **L. Lehtonen**, T. Tukiainen; Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland
Disclosure Block: **L. Lehtonen:** None.

Sex differences are common in human phenotypes, yet rarely discovered in genome-wide association studies (GWAS). Biobank-scale data has fuelled the search of sex differences in genetic associations but beyond waist-hip-ratio and testosterone the results have been largely negative.

To assess whether the lack of sex differences in genetic effects is due to the lack of power to discover these or simply because these do not exist, we examined the theoretical limits for the discovery of sex-biased associations in GWAS. We conducted power analyses for two strategies commonly applied for strata differences in GWAS: a genome-wide search for sex differences in genetic effects and a more focused approach assessing sex differences at variants reaching genome-wide significance in a sex-combined GWAS.

In the genome-wide strategy, with sample sizes similar to the UK Biobank (i.e. 150,000 per sex), the power to discover sex differences is minimal for anything but high effect variants with large differences. For example, to achieve 80 % power to detect differences at p=5e-8 for a variant explaining 1% of phenotype variance (R^2=1%) in one sex, R^2 should be 41% smaller in the other. For variants of R^2=0.1%, more common in GWAS, 80% power requires the variant to explain 13.4 times less variance in the other sex. Smaller differences and effects require substantial increases in sample size for well-powered assessments, e.g., N=440,000 per sex for R^2=0.1% and 25% difference in R^2 for 80% power. Expectedly, the focused approach has better power to discover variants with effects of concordant direction but different magnitude between the sexes, yet this approach misses variants with opposite effects. For such effects, the genome-wide search is well-powered, e.g., over 95 % power for all opposite direction effects when R2=0.1 % in the other sex with N=150,000 per sex.

These calculations demonstrate that even with the biobank-scale data currently available assessments of sex differences (or other two-strata differences) will likely be unsuccessful except for potential high-impact variants displaying considerable differences in effect magnitude. This result echoes the few sex differences discovered for most phenotypes, e.g., in the UK Biobank. Given the lack of power in these efforts, we can currently only exclude the large-scale existence of variants with large opposite effect sizes between the sexes. Unless manyfold sample sizes to these currently available can be collected, future analyses of sex differences should consider alternative strategies, such as genetic correlations and polygenic scores, that can reduce issues arising from the lack of power in individual variants.

PrgmNr 2094 - Challenges with X chromosome analyses and reporting in Genome-Wide Association Studies

View session detail

Author Block: **Z. Wang**[1], L. Sun[2,3], A. D. Paterson[4,3]; [1]Fudan Univ., Shanghai, China, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada, [4]Hosp for Sick Children, Toronto, ON, Canada
Disclosure Block: **Z. Wang:** None.
Little has improved regarding the analysis and reporting of X-chromosome variants from genome-wide association studies (GWAS) in the eight years since Wise et al., (2013) brought the eXclusion of the X-chromosome to the attention of the community. Using the EBI-NHGRI GWAS catalog (date downloaded 2020-03-08) we identified studies that reported genome-wide significant loci on the X-chromosome, and then extracted details from each. Out of 3869 studies in the catalog (male-only studies excluded), 195 reported a total of 564 loci on the X-chromosome, drastically fewer than 1308 studies reporting 5593 loci on chromosome 7, which has similar size to the X-chromosome.
Limitations of the X-chromosome analyses includes that most used methods designed for the autosomes to the X-chromosome: with sex as a covariate and additive coding of genotype; jointly analysing males and females; presumably with 0/2 coding of males assuming random X-chromosome inactivation. Rarely are sex-specific analyses reported, or sex differences in trait prevalence/trait distribution provided.
One study of refractive error identified two different loci in the pseudo-autosomal region PAR1 (PMID: 29808027) but did not describe how the analysis was performed and did not report sex-stratified results. Another study reported variants in the controversial PAR3 region to be significantly associated with ANCA-associated vasculitis, but it did not provide the sex distribution of the controls (PMID: 22808956) potentially making the association subject to confounding by sex. Only one locus has been identified in PAR2 (rs306890, associated with BMI and lipids; PMID:29507422; 30108127) but sex-specificity of the associations were not reported. Recent work (Chen et al. 2021; Song et al., 2021) has shown that genetic effect size estimated from sex-combined analyses are not interpretable.
Despite the major success of GWAS, the X-chromosome continues to be ignored or analyzed and reported in a suboptimal fashion. In addition to association testing, incorporating X-chromosome variants into polygenic risk scores poses additional analytic challenges.

PrgmNr 2095 - Comprehensive genome-wide association study of different forms of hernia identifies more than 80 associated loci

View session detail

Author Block: **F. Geller**[1], J. Fadista[2], L. Skotte[1], J. Karjalainen[3], E. Abner[4], E. SÃ¸rensen[5], H. Ullum[1], T. Werge[6], iPSYCH-group, T. Esko[7], L. Milani[4], A. Palotie[8], M. J. Daly[8], FinnGen, M. Melbye[9], B. Feenstra[1]; [1]Statens Serum Inst., Copenhagen, Denmark, [2]Statens Serum Inst., Copenhagen, Copenhagen, Denmark, [3]Boston, MA, [4]Univ. of Tartu, Tartu, Estonia, [5]RigsHosp.et, Copenhagen, Denmark, [6]Univ. of Copenhagen, Copenhagen, Denmark, [7]Univ. of Tartu, Tartu, Tartumaa, Estonia, [8]Massachusetts Gen. Hosp., Boston, MA, [9]Stanford Univ. Sch. of Med., Stanford, CA

Disclosure Block: **F. Geller:** None.

Introduction: A hernia is an outpouching of tissue through a preformed or secondarily established fissure. Epidemiologic studies suggest shared genetic etiology between different forms of hernia. Methods: Using data from the UK Biobank, we investigated five different forms of hernia classified with ICD 10 codes as inguinal (K40, 28,707 cases), diaphragmatic (K44, 31,193 cases), umbilical (K42, 6,402 cases), femoral (K41, 1,049 cases) and ventral hernia (K43, 4,644 cases). We also studied two combined groups: any hernia (65,492 cases) and any hernia excluding individuals who only had diaphragmatic hernia (38,762 cases). Results: We performed GWAS scans on the 7 described hernia groups, where each scan resulted in at least two associated loci. A replication study for inguinal hernia with almost 24,000 cases from four study groups confirmed 84 variants. Summing up all scans, we identified 114 independently associated variants in 81 loci (P -8). With MultiPhen we assessed a set of associated hernia types for each variant (P ADAMTS6, *ADAMTS16*, *ADAMTSL3*, *ELN*, *LOX*, *FBLN2* and *MFAP4*) or in the transforming growth factor Î² family (*TGFB2, BMP5*, *BMP7*, and *LTBP1*). Further evidence was derived by an inguinal hernia specific expression analysis with GeneNetwork resulting in 9 associated genes, 8 of which were in loci reaching genome-wide significance (*ADAMTS6*, *DNAJC27*, *EFEMP1*, *FBLN2*, *HAND2*, *LOX*, *MFAP4*). Conclusion: Overall, our study provides valuable insight into the genetics of herniation, identifying loci relevant for individual and multiple forms of hernia. The large number of associated loci for inguinal hernia point to several genes relevant for connective tissue homeostasis and general cell processes, which could be helpful to define genetic risk profiles or develop biomarkers indicating an increased vulnerability.

PrgmNr 2096 - Deep integrative models for large-scale human genomics

View session detail

Author Block: **A. Sigurdsson**[1], D. Westergaard[1], O. Winther[2,3], O. Lund[4], S. Brunak[1], B. Vilhjalmsson[5,6,7], S. Rasmussen[1]; [1]Novo Nordisk Fndn. Ctr. for Protein Res., Univ. of Copenhagen, Copenhagen, Denmark, [2]Bioinformatics Ctr., Dept. of Biology, Univ. of Copenhagen, Copenhagen, Denmark, [3]Section for Cognitive Systems, Dept. of Applied Mathematics and Computer Sci., Technical Univ. of Denmark, Kongens Lyngby, Denmark, [4]Danish Natl. Genome Ctr., Copenhagen, Denmark, [5]Aarhus Univ., Aarhus, Denmark, [6]Natl. Ctr. for Register-Based Res. (NCRR), Aarhus Univ., Aarhus, Denmark, [7]Lundbeck Fndn. Initiative for Integrative Psychiatric Res. (iPSYCH), Aarhus, Denmark
Disclosure Block: **A. Sigurdsson:** None.
Polygenic risk scores (PRSs) are expected to play a critical role in achieving precision medicine. PRS predictors are generally based on linear models using either individual-level data or summary statistics. However, these predictors generally do not capture nonlinear relationships, epistatic interactions, or genotype-environment interactions. Here, we develop a deep learning framework (EIR) for PRS prediction which includes a model, genome-local-net (GLN), specifically designed for large scale genomics data. The framework supports multi-task (MT) learning, automatic integration of clinical and biochemical data and model explainability. We apply EIR to predict 338 health outcomes, representing a wide range of different diseases in the UK biobank. We find that the EIR generally outperforms LASSO prediction, and in particular for autoimmune disease which were previously shown to have significant epistatic interactions. We showcase the flexibility of the framework by training one MT model to predict the 338 diseases simultaneously. Furthermore, we find that incorporating clinical measurement data for PRSs improves performance for virtually all (93%) diseases considered (ROC-AUC improvement up to 0.36) and that including genotype data provides better model calibration compared to measurements alone. We use the framework to analyse what our models learn and find that they learn relevant disease variants and clinical measurements. EIR is open source and available at https://github.com/arnor-sigurdsson/EIR.

PrgmNr 2097 - Genome-wide association study of chronotype in Japanese direct-to-consumer genetic testing data

View session detail

Author Block: **S. Nogawa**[1], Y. Tahara[2], M. Chijiiwa[1], K. Saito[1], S. Shibata[2]; [1]Genequest Inc., Tokyo, Japan, [2]Waseda Univ., Tokyo, Japan
Disclosure Block: **S. Nogawa:** Salary/Employment; Genequest Inc..
[Background]
Chronotype refers to the trait of morning or evening type. Our sleep phase should fit our chronotypes to achieve good health. A study in European population has reported that many SNPs are associated to chronotypes. However, a genome-wide association study (GWAS) of chronotype has not been performed in East Asian populations. In this study, we performed a chronotype GWAS for the Japanese population to identify East Asian-specific SNPs associated to chronotypes.
[Methods]
Study subjects were about 12,000 consumers of a direct-to-consumer genetic testing service. Their genetic data was analyzed by genotyping arrays. Chronotype was calculated as a sleep-corrected midpoint of sleep in free days value (MSFsc) based on the self-reported questionnaire. MSFsc value enable the measurement of individual chronotypes accurately because of a continuous variable.
[Results]
The chronotype GWAS identified rs11066015 (*p-value*=7.51E-17) in 12q24 locus; 12q24 locus are East Asian-specific polymorphisms and include *ALDH2* related to alcohol metabolism. This association may show that alcohol habit has a significant impact on chronotype. We also compare this GWAS result with European GWAS data. This study was the first chronotype GWAS in East Asian population and provided new insights into circadian rhythm research.

PrgmNr 2098 - Genome-wide association study of varicose veins identifies a protective low-frequency missense variant in GJD3 enriched in Finnish population and highlights connexins as potential therapeutic targets

View session detail

Author Block: **S. Hassan**[1], E. Saarentaus[2], E. Vartiainen[1], A. Palotie[3], J. Karjalainen[1], M. Kurki[4], S. Ripatti[5], T. Tukiainen[2]; [1]Univ. of Helsinki, Helsinki, Finland, [2]Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland, [3]Massachusetts Gen. Hosp., Boston, MA, [4]Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Inst. for Molecular Med. Finland FIMM, Helsinki, Finland
Disclosure Block: **S. Hassan:** None.
Varicose veins (VV) is the most common manifestation of chronic venous disease, increasing risk for severe vascular conditions, like venous thromboembolism (VTE). In a genome-wide association study of VV in Finnish population (17027 VV cases, 190028 controls), we identify 45 significantly associated loci (p10x enriched in Finns. A low-frequency protective missense variant in GJD3 (p.Pro59Thr; p=1.03e-14, OR(95%CI)= 0.62(0.55-0.70), member of the connexin gene family is exclusively associated with VV in a phenome-wide scan, pointing to the potential of connexin-modulating therapeutic strategies for VV. Our results provide novel insights into VV aetiopathology and highlight the power of isolated populations like Finland to discover genetic variants informing therapeutic development.

PrgmNr 2099 - Genome-wide phenome-wide association study for important diseases & quantitative traits based on the Taiwan Biobank data

View session detail

Author Block: **H-C. Yang**[1], C-W. Chen[1], J-H. Wang[1], H-C. Liao[1], T-C. Chan[2], J-S. Hwang[1], M-W. Su[3], T-C. Lee[3], C-W. Lin[3], C-h. Chen[1]; [1]Inst. of Statistical Sci., Academia Sinica, Taipei, Taiwan, [2]Res. Ctr. for Humanities and Social Sci., Academia Sinica, Taipei, Taiwan, [3]Inst. of BioMed. Sci., Academia Sinica, Taipei, Taiwan

Disclosure Block: **H. Yang:** None.

Genome-wide phenome-wide association study provides a comprehensive search for pathogenic germline variants and novel candidate predisposition genes with pathways for diseases and quantitative traits (QTs). In this study, we conducted a genome-wide phenome-wide association study for 28 self-reported diseases and 37 QTs based on the Taiwan Biobank data. After data quality control, 21,043 out of 24,000 individuals and 600,066 out of 646,951 single nucleotide polymorphisms (SNPs) remained. SNPs and the clusters of SNPs with linkage disequilibrium susceptible for cardiomyopathy, diabetes, gout and hyperlipidemia are identified; 3,037 genetic association signals for QTs are also found. We constructed polygenic risk scores (PRS) for the studied diseases and QTs. The results of strong association and high prediction for QTs based on the constructed PRS are found. A SNP-based heritability analysis reveals a high genetic heritability of QT, but the heritability estimates for most of the studied diseases are limited. We also conducted the genome-wide gene-environmental interaction studies that account for the particulate matter (PM2.5). The analysis identifies a number of SNP-PM2.5 interactions for several diseases such as osteoporosis, peptic ulcer, postpartum depression as well as for several QTs such as fasting blood glucose and serum glutamic-pyruvic transaminase. Genetic network analysis for complex disorders and QTs based on the identified SNPs and PRSs establishes the disease and QT clusters and their connections. A knowledge database for the identified association signals is established. This study provides an evidence-based perspective for the complex mechanism of genetic etiology.

PrgmNr 2100 - Genotype-by-environment interactions for chronic back pain

View session detail

Author Block: **I. Kuznetsov**[1], Y. A. Tsepilov[2], Y. S. Aulchenko[3]; [1]Skolkovo Inst. of Sci. and Technology, Moscow, Russian Federation, [2]Novosibirsk State Univ., Novosibirsk, [3]PolyOmica, Maastricht, Netherlands
Disclosure Block: **I. Kuznetsov:** None.
Introduction Back pain (BP) is a common debilitating condition with a lifetime prevalence of 40% and a large social and economic impact. Known clinical risk factors for BP include age, female sex, and raised body mass index (BMI). In the majority of BP episodes, the symptoms are transient; however, about 10% of those experiencing acute BP develop a chronic condition. Studies of chronic back pain demonstrated this is a complex trait under highly polygenic control (Freidin et al., 2019). Genetic variation in response to the environment is fundamental in biology and has been described as genotype-by-environment interaction (GxE). Taking into account relevant environmental specificity of genetic effects can improve the fidelity and statistical power of genome-wide association analyses and raise implications for treatment and preventive measures for diseases (Shin and Lee, 2020). In this work, we studied the role of GxE interactions in the control of chronic BP.
Materials and Methods We analyzed ~335,000 unrelated white UK Biobank participants of British ancestry. The trait of interest was â□□Back pain for 3+ monthsâ□□, adjusted for the potential confounders. Age, sex, BMI, education level, and income were tested as covariates interacting with genotype. To estimate heritability explained by GxE we used whole-genome GxE analysis as implemented in GxEsum (Shin and Lee, 2020) - a method based on GWAS summary statistics, using an extension of the LD score regression approach.
Results We estimated the heritability of chronic BP explained by interactions of genotype with age, sex, BMI, education, and income. All estimates did not significantly differ from zero. The highest estimated h2 = 0.59% (SE=0.33%) was for sex.
Discussion GxE interactions for chronic back pain were not found with the whole-genome approach used here. This observation can be partly explained by modest effect sizes and insufficient sample size. The method we used is not sensitive to scale effects which could be the reason for observed GxE interactions in previous studies.
Conclusion No evidence for interactions of genotype by common environmental risk factors for chronic back pain was found.
Acknowledgements This research has been conducted using the UK Biobank Resource (project # 41601, â□□Non-additive effects in control of complex human traitsâ□□).
References Freidin, M. B. et al. (2019) â□□Insight into the genetic architecture of back pain and its risk factors from a study of 509,000 individualsâ□□, Pain. Shin, J. and Lee, S. H. (2020) â□□GxEsum: genotype-by-environment interaction model based on summary statisticsâ□□, BioRxiv.

PrgmNr 2101 - Hierarchical topic modelling enables genetic analysis of electronic healthcare record data

View session detail

Author Block: **y. zhang**[1], G. Lunter[2,3], G. McVean[4]; [1]Big Data Inst., Univ. of Oxford, Oxford, United Kingdom, [2]Univ. Med. Ctr. Groningen, Dept. of Epidemiology, Univ. of Groningen, Groningen, Netherlands, [3]Weatherall Inst. of Molecular Med., Univ. of Oxford, Oxford, United Kingdom, [4]Univ. of Oxford, Oxford, United Kingdom
Disclosure Block: **Y. zhang:** None.
Genome-wide association studies have identified thousands of loci across the genome that influence risks for diseases, many of which have pleiotropic effects. Conversely, at the phenotypic level, certain diseases are frequently found to co-occur in individuals (multi-morbidities). We have developed an approach to unite these two perspectives. By incorporating hierarchical medical ontologies with an adapted version of latent Dirichlet allocation (LDA), we have developed â□□treeLDAâ□□ to model patterns of comorbidity. Such an approach both advances our understanding of relationships between diseases and allows us to analyse genetic association in a lower dimensional â□□topicâ□□ space, which may be more reflective of the underlying causal pathways for diseases.
We show, using simulation and empirical analysis of hospital episode statistic (HES) data from the UK Biobank, coded using the ICD10 ontology, that treeLDA identifies disease topics that align with current medical understanding, outperforms existing topic modelling methods in specific situations, notably when training data is small and/or individuals often have multiple disease topics, and provides us with additional power for genetic discovery. For example, at genome-wide significance we identified 2086 loci associated with disease topics (allowing up to 32 distinct topics), compared to 815 loci associated with single disease codes. We also show that common disease codes are often associated with multiple topics, which can improve genetic risk prediction. Conversely, we identified 126 loci associated with a â□□healthyâ□□ topic characterised by an absence of active disease codes.
Our topic modelling approach provides a data driven approach to model multi-morbidity that can address the challenges of sparseness and heterogeneous presentation and recording practice inherent in electronic healthcare record data. Finally, we show that modelling topics at different levels of resolution (i.e. numbers of topics), combined with genetic analysis, provides a means of identifying the level of phenotypic complexity at which genetic influences act.

PrgmNr 2102 - Machine-learning SNP-based prediction for Primary Biliary Cholangitis: A proof-of-concept study

View session detail

Author Block: **A. Gerussi**[1], D. Verda[2], L. Cristoferi[1], E. Paraboschi[3], G. Mulinacci[1], M. Muselli[2], M. Carbone[1], P. Invernizzi[1], R. Asselta[3]; [1]Univ. of Milano-Bicocca, Monza, Italy, [2]Rulex, Genova, Italy, [3]Humanitas Univ., Pieve Emanuele, Italy
Disclosure Block: **A. Gerussi:** None.
Background: Polygenic risk scores (PRS) are commonly used in risk prediction models for complex traits, but suffer from several limitations. Only one PRS was developed in Primary Biliary Cholangitis (PBC), based on alleles evidenced by genome-wide association studies (GWAS), but published before many novel variants were discovered through huge meta-analyses. Here, we aimed to evaluate the feasibility and accuracy of a Machine learning (ML)-based model for risk prediction, which has several methodological advantages over PRS.
Methods: Genome-wide significant variants identified in subjects of European ancestry in the recently released 2nd international meta-analysis of GWAS in PBC were used as input data. Quality-checked, individual genomic data from Italian patients and controls were used. The main analytic steps were the following: import of genotype and phenotype data, feature ranking, feature selection, supervised classification of phenotype by genotype, generation of if-then rules for disease prediction by logic learning machine (LLM). Ten-fold cross-validation was performed for validation. RulexÂ© software was used to build the ML model, while PLINK 1.9 was used for extraction of variants of interest and preliminary QC steps. A cost of function to penalize the model and improve accuracy of PBC prediction was added.
Results: 1345 individuals were included in the study, 444 were PBC cases and 901 healthy controls. 515 were males and 830 females. 105150 variants were imported in RulexÂ© and, after filtering missing values, 41899 variants were available for the analysis. Several configurations of parameters related to feature selection were simulated. The time for a complete run (from import to rules generation) is 2.30 hours on average on a 64gb working station. The final model had an accuracy of 0.73, a Youdenâ□□s value of 0.32, a sensitivity of 0.44, a specificity of 0.88, a positive predictive value of 0.64 and a negative predictive value of 0.76. Five rules where generated with a covering > 20%. The rule with the highest covering, (44%) had an error of 16% and seven conditions (female sex, and gene variants associated with genes *CD80*, *WDFY4* and *CLEC16A*). The rule with the lowest number of conditions had a covering of 27% and an error of 6% and was based on one variant within the HLA region and female sex.
Conclusion: This study represents the first illustration of a successful analysis of GWAS data with ML for PBC. ML is computationally feasible and generates accurate biological information that can be leveraged for disease prediction. Future steps will include external validation and comparison with a PBC-specific PRS based on the same set of variants.

PrgmNr 2103 - Phenome-wide association study of deleterious variant (s_{het}) burden using exome sequencing of 394,694 UK Biobank participants

View session detail

Author Block: **K. J. Carss**[1], O. Burren[1], Q. Wang[1], F. Hu[1], D. Vitsios[1], E. J. Gardner[2], M. E. Hurles[3], S. Petrovski[1]; [1]AstraZeneca, Melbourn, United Kingdom, [2]Wellcome Sanger Inst., Hinxton, United Kingdom, [3]Wellcome Sanger Inst., Cambridge, United Kingdom
Disclosure Block: **K.J. Carss:** Salary/Employment; AstraZeneca.
The selection coefficient (s_{het}) burden score was introduced by Gardner and colleagues as a statistic to reflect the exome-wide burden of deleterious variants carried by individuals.[1] They showed s_{het} burden to be negatively correlated with fecundity in humans. In this study we tested the hypothesis that a modified version of s_{het} burden score could be associated with a wider range of human traits. We calculated an s_{het} burden score for each of 394,694 UK Biobank participants by focusing on rare (MAFhet burden score against ~16500 phenotypes including binary clinical endpoints as well as quantitative traits, correcting for age and sex.
We identified a total of 132 binary and 61 quantitative traits that achieved a p-8 association with our modified version of the s_{het} burden score. These include the previously reported association between PTV burden and mental and behavioural disorders ($p=8.7x10^{-22}$) as well as associations with leukaemia, respiratory and metabolic disorders.
Future iterations will incorporate additional variant types such as deletions called using microarrays as well as deleterious-predicted missense variants occurring in constrained regions of the human exome. Already, however, the current results yield known and novel insights into the genetic architecture of many complex traitsâ□□including many that have not been thoroughly studied using sequencing technologiesâ□□especially those traits with high locus heterogeneity and where the genetic signals are driven by variants too rare to be detected by individual variant analyses.
1) Gardner, E. J. *et al.* Sex-biased reduction in reproductive success drives selective constraint on human genes. *bioRxiv* 2020.05.26.116111 (2020). doi:10.1101/2020.05.26.116111

PrgmNr 2104 - Polygenic Risk Score Estimation in North-Western Russian Population

View session detail

Author Block: **V. Rezapova**[1,2], N. Kolosov[1,2], O. Rotar[1,2], O. Freylikhman[1], A. Loboda[1,2], O. Melnik[1,2], A. Sergushichev[2], A. Gnirke[3], C. Stevens[3], A. Meissner[4], A. Kostareva[1,2], A. Konradi[1,2], M. Daly[3,5,6], M. Artomov[1,3,6,2]; [1]Almazov Natl. Med. Res. Ctr., Saint-Petersburg, Russian Federation, [2]ITMO Univ., Saint-Petersburg, Russian Federation, [3]Broad Inst., Cambridge, MA, [4]Max Planck Inst. for Molecular Genetics, Berlin, Germany, [5]Finnish Inst. for Molecular Med., Helsinki, Finland, [6]Massachusetts Gen. Hosp., Boston, MA

Disclosure Block: **V. Rezapova:** None.

Over the last decade, genome-wide association studies (GWAS) have discovered a substantial number of associated variants for many complex traits. However, even within European-centered GWAS data, there are local subpopulations significantly under-represented in these studies. For example, Russians, being one of the largest ethnic groups among the Europeans, remained significantly under-represented in GWAS for years.We used a pilot genotyping cohort of 239 individuals from Saint-Petersburg to investigate phenotypic variance explained by polygenic risk scores (PRS) for 11 phenotypes. We used UK biobank (UKBB) GWAS summary statistics for corresponding phenotypes and selected optimal p-value thresholds for maximizing R^2 for PRS.Several strategies for PRS calculation were tested, including effects of genotype imputation and usage of sex-specific GWAS summary statistics. In addition, we compared R^2 estimates for polygenic risk models in a scenario when UKBB GWAS summary statistics is applied to target data from UKBB itself or Biobank Japan (BBJ). The best utility of UKBB GWAS was observed for UKBB participants, with predictive value for Russian-descent individuals taking an intermediate place between UKBB and BBJ.

PrgmNr 2105 - Quantifying concordant genetic effects of de novo mutations on multiple disorders

View session detail

Author Block: **H. GUO**[1], L. Hou[2], Y. Shi[3], S. Jin[4], X. Zeng[5], B. Li[6], R. Lifton[7], M. Brueckner[7], H. Zhao[8], Q. Lu[9]; [1]Tsinghua Univ., Beijing, China, [2]Beijing, China, [3]New Haven, CT, [4]Washington Univ. Sch. of Med., St. Louis, MO, [5]The Rockefeller Univ., New York, NY, [6]Yale Sch. of Publ. Hlth., New Haven, CT, [7]Yale Univ., New Haven, CT, [8]Yale Univ. Sch. of Publ. Hlth., New Haven, CT, [9]Univ. of Wisconsin-Madison, Madison, WI

Disclosure Block: **H. Guo:** None.

Exome sequencing on tens of thousands of parent-proband trios has identified numerous deleterious *de novo* mutations (DNMs) and implicated risk genes for many disorders. Recent studies have suggested shared genes and pathways are enriched for DNMs across multiple disorders. However, existing analytic strategies only focus on genes that reach statistical significance for multiple disorders and require large trio samples in each study. As a result, these methods are not able to characterize the full landscape of genetic sharing due to polygenicity and incomplete penetrance. In this work, we introduce EncoreDNM (**En**richment **cor**relation **e**stimator for ***D****e* ***N****ovo* **M**utations), a novel statistical framework to quantify shared genetic effects between two disorders characterized by concordant enrichment of DNMs in the exome. EncoreDNM makes use of exome-wide, summary-level DNM data, including genes that do not reach statistical significance in single-disorder analysis, to evaluate the overall and annotation-partitioned genetic sharing between two disorders. Applying EncoreDNM to DNM data of nine disorders, we identified abundant pairwise enrichment correlations, especially in genes intolerant to pathogenic mutations and genes highly expressed in fetal tissues. These results suggest that EncoreDNM improves current analytic approaches and may have broad applications in DNM studies.

PrgmNr 2106 - Sources of Inequality at Birth: The Interplay Between Genes and Parental Socioeconomic Status

View session detail

Author Block: **A. Marees**[1], P. Biroli[2], T. Galama[3], P. Arce[2], J. Vollen[2]; [1]Vrije Univ.it Amsterdam, Amsterdam, Netherlands, [2]Univ. of Zurich, Zurich, Switzerland, [3]Univ. of Southern California, Los Angeles, CA

Disclosure Block: **A. Marees:** None.

The start of a humanâ□□s life can be characterized by two lotteries: that of your genes (nature) and that of the family you were born into (nurture). These lotteries set in motion a trajectory, from birth onward, in health and human-capital formation. Genetic factors are implicated in virtually any human trait, and early-life experiences strongly influence long-term outcomes, with children born into high socio-economic status (SES) families generally in better health and of higher SES in later life. Leveraging three longitudinal social-science data sets, the Health and Retirement Study (HRS), the Wisconsin Longitudinal Study (WLS), and the English Longitudinal Study of Aging (ELSA), containing rich genotypic and phenotypic information, we systematically analyze the relationship between an individual's genetic predisposition towards a trait (genotype), their actual trait in adulthood (phenotype), and the social and economic status of the families they grew up in. We proxy the individual's genetic predisposition to a trait by a polygenic score (PGS) and the SES of the families they were born into by a latent factor of parental SES, constructed from parental education and the fatherâ□□s (former) occupational status. We then investigate how genetic predispositions, socio-economic background, and their interaction contribute to later-life outcomes, across a range of forty-five socioeconomic, anthropometric, health, behavioral, and personality traits. We find strong genetic and socio-economic associations but no evidence for sizeable interactions between them. We discuss several possible implications of this result.

PrgmNr 2107 - Systematic comparison of family history and polygenic risk across 27 common diseases

View session detail

Author Block: **N. Mars**[1], J. V. Lindbohm[2,3], P. della Briotta Parolo[1], E. WidÃ©n[1], J. Kaprio[1,2], A. Palotie[1,4,5], FinnGen, S. Ripatti[1,6]; [1]Inst. for Molecular Med. Finland, HiLIFE, Univ. of Helsinki, Helsinki, Finland, [2]Clinicum, Dept. of Publ. Hlth., Univ. of Helsinki, Helsinki, Finland, [3]Dept. of Epidemiology and Publ. Hlth., Univ. Coll. London, London, United Kingdom, [4]Analytic and Translational Genetics Unit, Dept. of Med., Massachusetts Gen. Hosp., Boston, MA, [5]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Broad Inst. of MIT and Harvard, Cambridge, MA
Disclosure Block: **N. Mars:** None.
Family history is the standard measure of inherited disease susceptibility in clinical practice, while polygenic risk scores (PRS) have recently been shown to capture efficiently personal genetic risks in many diseases. No studies have systematically compared how they overlap and complement each other in a wide range of common complex diseases.
Here we leverage family relationships, information about parental causes of death, and genome-wide genotyping of 306,418 individuals within the FinnGen study, to systematically examine the interplay of first and second-degree family history of diseases, parental causes of death, and genome-wide PRSs. With up to 50 years of follow-up within nationwide health registries, we study 27 diseases, covering a large proportion of the burden of non-communicable diseases in adults. The 27 genome-wide PRSs were constructed using PRS-CS.
The effects of PRS and family history were largely independent and provided complementary information for risk prediction. Both high PRS and family history were strongly associated with respective diseases, but a high PRS (>90th percentile) conferred larger risks than family history in cardiometabolic diseases and in the three common cancers studied. Among 39,444 individuals with a first-degree relative in the dataset, the impact of adjusting the PRS with first-degree family history was small (mean decrease in betas -3.4%, s.d. 1.8%). Similarly, the decrease was small adjusting the effect of first-degree family history with the PRS (-4.9%, s.d. 3.3%). The decreases were even smaller for second-degree family history (N=47,154), and for parental causes of death (N=227,982). In 23/27 diseases, PRS improved risk discrimination beyond first-degree family history, with largest increases in rheumatoid arthritis, asthma, and prostate and breast cancer.
PRSs were effective also for risk stratification among individuals with positive family history. In most diseases, a positive family history with a high PRS was associated with a considerably elevated risk, whereas a low PRS compensated completely for the risk implied by a positive family history. For instance, in type 2 diabetes, individuals with positive family history and high (>90th percentile), average (33-90th) or low (This study provides a catalogue of risk estimates and prediction accuracy for both family history of disease and PRSs. We demonstrate that these metrics are largely independent and complementary measures of familial susceptibility.

PrgmNr 2108 - Systematic comparison of performance of over 50 PRS across ancestries for 14 diseases

View session detail

Author Block: **V. Plagnol**, S. Selzam, W. Tarran, R. Moore, D. Wells, F. Riveros-Mckay, P. Donnelly, G. McVean, M. E. Weale; Genomics plc, Oxford, United Kingdom
Disclosure Block: **V. Plagnol:** Salary/Employment; Genomics plc.
There has been intense recent interest in the performance of PRSs across ancestry groups, in particular in the extent of performance deterioration in non-European individuals. This is particularly pressing as the technology moves closer to use in healthcare.
However, like-for-like comparison of reported performances is often difficult, for example because of differences in cohort recruitment, phenotype definition, population characteristics or use of covariates such as age/sex in reported metrics. Even if the same evaluation cohort was used, many of these interpretation issues remain. To overcome these issues we established a uniform PRS comparison pipeline. This leverages the UK Biobank and several ethnically diverse cohorts (Multi-Ethnic Cohort, BioMe, ARIC, MESA, PAGE) to enable direct comparison between over 50 published and internally developed PRS models for 14 complex diseases in four ancestry groups (European, South Asian, East Asian, African), where sufficient case numbers were available for analysis.
Maximum AUC values were observed for individuals of European ancestries, followed by East Asians (average AUC reduction 2.36%), South Asians (average AUC reduction 4.51%) and Africans (average AUC reduction 10.8%). A few publicly available PRS are optimised for a specific ancestry, but for these we did not observe a predictive boost in the target ancestry. Genome-wide PRS including millions of variants were systematically more effective than sparser models, even when European linkage disequilibrium (LD) was applied to non-European ancestries (e.g. cross ancestry AUC boost of 8.2% in breast cancer). Inclusion of functional annotations yielded slight improvements (average ÎAUC 0.005) across ancestries and diseases. Internal optimization using population specific LD maps and bespoke cross-ancestry methodologies generated a detectable increase in the target ancestries, provided that a training set with an effective sample size â¥20% of the European set was available. For example, type 2 diabetes prediction improved across all ancestries, for a minimum of 10% for African ancestries (ÎAUC 0.06) to a maximum of 16% for East Asian ancestries (ÎAUC 0.09).
Systematic comparisons of PRS models across ancestry groups shows that the expected deterioration of performance was relatively limited between European and East/South Asian ancestries, but much more marked in individuals with African ancestries. Choice of PRS methodology substantially impacted predictive performance, suggesting that methodological developments as well as increasing data availability from non-European studies, will improve PRS performance across ancestries.

PrgmNr 2109 - Web-based, participant-driven research platform discovered novel genetic associations for stress-related complaints and food-related behaviors in Japanese population

View session detail

Author Block: **S. Ishida**[1], M. Tanaka[1], S. Sassa[1], K. Katayama[2], S. Miyano[2], S. Imoto[2]; [1]DeNA Life Sci., Inc., Tokyo, Japan, [2]Div. of Hlth.Med. Intelligence, Human Genome Ctr., The Inst. of Med. Sci., The Univ. of Tokyo, Tokyo, Japan
Disclosure Block: **S. Ishida:** Salary/Employment; DeNA Life Science, Inc..
Genome-wide association studies (GWAS) have identified thousands of loci linked to human complex diseases and traits mainly in European ancestry populations. However, non-European studies are needed to fully understand genetic architectures across diverse populations. Since 2014, we have provided the direct-to-consumer genetic testing service MYCODE for the Japanese population, which is an Asian population with a more genetically homogeneous community than other populations such as Europeans. This homogeneity is a suitable feature for GWAS. To date, the service has reached more than 100,000 users. We have also created a research platform, MYCODE Research, where 90% of the users give consent for use of their genome-wide SNP data and health-related information for research. We used this web-based research framework to enable rapid recruitment of research participants and collect large volumes of data, thereby accumulating over 4 million self-reported phenotypic data. Users can participate in research of interest, and receive personalized feedback, such as analyzed results or lifestyle advice. We performed GWASs against over 50 self-reported traits such as biochemical and psychological traits, using up to 53,000 Japanese individuals. A total of 141 genome-wide significant associations were obtained, including 134 known genetic associations. Importantly, 12 out of them were previously reported from traditional medical record-based studies, demonstrating the reliability of our self-reported phenotypic data for population genomics. We newly identified 7 genetic associations for stress-related complaints and food-related behaviors, and 5 of them were located on or close to gene regions. The alcohol-related *ALDH2* gene variant showed pleiotropic associations. The allele with a decreasing effect on alcohol consumption increased fruit consumption and snack purchasing behavior, while decreased difficulty staying asleep.These association signals disappeared after adjusting for drinking behavior, suggesting that the *ALDH2* genotype modifies these phenotypes via alcohol consumption. The *HCN2* gene locus was associated with irritability, and this locus included a missense variant (rs113534512) associated with febrile seizure and generalized epilepsy. The 6q25.2 locus was associated with difficulty falling asleep, and eQTL datasets showed a significant correlation between the lead SNP, rs533737, and expression of *MTRF1L*, whose variants were associated with chronotype. These results suggested shared biological mechanisms involving *MTRF1L* between chronotype and sleep-related complaints.

PrgmNr 2110 - Epigenome-wide association study of DNA methylation in CD4- and CD8-positive T cells in narcolepsy

View session detail

Author Block: **M. Shimada**[1], T. Miyagawa[2], K. Yoshida[3], K. Tokunaga[4], M. Honda[5]; [1]Natl. Ctr. for Global Hlth.and Med. (NCGM), Tokyo, Japan, [2]Tokyo Metropolitan Inst. of Med. Sci., Tokyo, Tokyo, Japan, [3]The Univ. of Tokyo, Tokyo, Japan, [4]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [5]Tokyo Metropolitan Inst. of Med. Sci., Tokyo, Japan
Disclosure Block: **M. Shimada:** None.
Narcolepsy type 1 (NT1) is a hypersomnia characterized by excessive daytime sleepiness and cataplexy. NT1 is caused by a loss of orexin-producing cells in the posterior hypothalamus, however, why the neurodegeneration occurs is still unclear. Previous studies suggested a strong association with *HLA-DQB1*06:02* and the involvement of other immune-related genes such as *CCR1/CCR3*, indicating that abnormalities in the immune system might be involved in disease development. Recently, autoreactive CD4+ and CD8+ T cells specific for orexin and proteins specifically expressed by orexin-producing cells respectively were reported to be associated with narcoleptic etiology and the importance of these T cells in narcolepsy has been suggested. In this study, we performed a microarray-based epigenome-wide association study (EWAS) of DNA methylation for NT1 using CD4+ and CD8+ T cells separated from PBMC of the same origin using magnetic beads (Case; N = 28, Control; N = 28). With proper filtering and normalization, we obtained data on methylation rates for 746,913 CpG sites. PCA using the methylation rate of all the sites found that the variance of PC1 and PC2 could be largely explained by the methylation rate of T-bet and GATA3-binding sites, respectively. Then we examined whether the ratio of subtypes of the CD4+ or CD8+ T cells differed between NT1 and control subjects. We performed estimation of cell components with both reference-based and reference-free analytical methods and found that the ratio of these T cell subtypes did not differ greatly between NT1 and control subjects. Finally we performed EWASs of CD4+ and CD8+ T cells and replication EWASs (Case; N = 14, Control; N = 14) and searched differentially methylated regions (DMRs) of which associations with NT1 were replicated. As a result, 15 DMRs and 5 DMRs were identified in the CD4+ and CD8+ T cells respectively. Most DMRs were common in both T cells; the associations in CD8+ T cells tended to be weaker than those in CD4+ T cells, and only 5 DMRs were identified both in the initial and replication EWASs, however, many of the DMRs detected in CD4+ T cell analysis were detected in either the initial or replication EWAS of CD8+ T cells. Although gene enrichment analysis using DMR-neighboring genes did not suggest any particular immune pathway, several DMRs in CD4+ T cells were located upstream of genes such as *CCL5*, *ALOX12,* and *EIF4G1* related to positive regulation of cell death. In particular, *CCL5* was detected in both CD4+ and CD8+ T cell analyses and is a chemokine-related gene that has been reported to be associated with NT1, suggesting that *CCL5* is an interesting candidate gene for further investigation.

PrgmNr 2111 - Functional characterisation of *GJB2 cis*-regulatory elements and WGS of heterozygous patients with NSHL

View session detail

Author Block: **A. LE NABEC**[1], M. COLLOBERT[2], C. LE MARECHAL[1,3], S. MOISAN[1,3], C. FEREC[1,3]; [1]Univ Brest - Inserm UMR 1078, Brest, France, [2]Univ Brest - Inserm UMR 1078, FRANCE, France, [3]CHRU Brest, Brest, France

Disclosure Block: **A. Le nabec:** None.

Three-dimensional chromatin organization plays a key role on gene expression. Gene regulation depends on *cis*-regulatory elements which can interact with gene promoter by chromatin loop. Alteration of chromatin architecture and/or *cis*-acting elements can lead to *enhanceropathies*. Several unelucidated nonsyndromic hearing loss and deafness 1 (*DFNB1)* cases carrying out only one heterozygous pathogenic mutation on Gap Junction Beta 2 (*GJB2*) gene, led to strongly suggest the presence of distant *cis*-regulation. Thanks to chromatin conformation study, we previously identified several *cis*-regulatory elements (CRE) which have enhancer action and silencer effect on *GJB2* expression. Analysis of CTCF binding allowed to purpose a *DFNB1* 3D looping model. We identified cooperative effects between enhancers. To confirm an endogenous enhancer activity, we develop CRISPR interference (CRISPRi), new approach for targeted silencing of transcription in human cells. We target *GJB2 cis*-acting elements with dCas9-KRAB. Preliminary results show a decrease of *GJB2* expression. Moreover, we develop 4C technique to analyse interactions of *GJB2* CREs and confirm a cooperative effect. Then, we focus on 10 patients with incomplete genotype. We realize a whole genome sequencing with HiSeq 4000 by IntegraGen Genomics. WGS analysis allowed to redress three genotype. Moreover, we realize functional assays to analyse a variant in *GJB2* promoter and continue analyses for the others genotypes. This work is supported by grants from the French foundation â□□La Fondation pour lâ□□Auditionâ□□, the â□□RÃ©gion Bretagneâ□□ and the association â□□GaÃ©tan SalaÃ¼nâ□□.

PrgmNr 2112 - Genome-wide Analysis of DNA methylation in 106 Schizophrenia family trios in Han Chinese

View session detail

Author Block: **L. Shen**; Shanghai Jiao Tong Univ., Shanghai, China
Disclosure Block: **L. Shen:** None.
Schizophrenia (SCZ) is a severe neurological disorder, affecting approximately 0.75% of population worldwide. It is well known that schizophrenia tends to run in family while both genetic and environmental factor contribute to its etiology. Overwhelming evidence suggested that alterations in DNA methylations occurred in SCZ patients. To investigate potential inheritable pattern of DNA methylation in SCZ family, we carried out a genome-wide analysis of DNA methylation of peripheral blood samples from 106 Chinese SCZ family trios. Genome-wide DNA methylations of CpG islands were quantified by Agilent 1Ã□244k Human Methylation Microarray. In this study, we proposed a loci inheritance frequency model to characterize differential methylated regions as SCZ biomarkers. Based on this model, 112 hypermethylated and 125 hypomethylated regions were identified. 121 hypermethylated and 139 hypomethylated genes were annotated. By functional enrichment analysis, hub gene networks indicated that many DMGs involved in Notch/HH/Wnt signaling, MAPK signaling, GPCR signaling, immune response signaling. Notably, the hypomethylated genes were significantly tissue-specific enriched in cerebral cortex and functionally enriched in nervous system development. The findings not only proved previous well-characterized SCZ risk genes but also provided novel candidate DMGs in SCZ. In sum, our results may further the understanding of altered DNA methylations in SCZ.

PrgmNr 2113 - Genome-wide DNA methylation profiling of leukocytes identifies CpG methylation signatures of treatment-resistant schizophrenia

View session detail

Author Block: **K-M. Lu**[1], S. Hsieh[2], S-H. Lin[1]; [1]Inst. of Clinical Med., Coll. of Med., Natl. Cheng Kung Univ., Tainan City, Taiwan, [2]Dept. of Psychology, Coll. of Social Sci., Natl. Cheng Kung Univ., Tainan City, Taiwan
Disclosure Block: **K. Lu:** None.
Background Treatment-resistant schizophrenia (TRS) is one of the schizophrenia subgroups, defined as a non-response to at least two trials of antipsychotic medication of adequate dose and duration. To establish epigenetic DNA methylation changes for treatment-resistant schizophrenia in the genome, methylation signatures have been tested to reflect how influences the brain functions and neuronal activities and identify potential mechanisms. We aimed to evaluate the discriminant abilities of DNA methylation probes between treatment-resistant schizophrenia and nontreatment-resistant schizophrenia. **Methods** In this study, 8 treatment-resistant schizophrenia patients and 8 nontreatment-resistant schizophrenia patients (ages ranging from 20 to 65 years old) were recruited. The blood samples of patients were collected and the resulting DNA samples were used for human MethylationEPIC 850K BeadChip scanning of global DNA methylation pattern and DNA methylation Î²-values at individual candidate CpG sites in the 850K arrays. And we conducted enrichment analysis by significant probes. **Results** Based on genome-wide scans of 8 treatment-resistant schizophrenia versus 8 nontreatment-resistant schizophrenia, 111 differentially methylated probes were identified at FDR30% differences DNA methylation Î²-values. These differentially methylated CG probes and the CpG island-associated genes were analyzed with gene Set Enrichment Analysis. Our results identified four genesets that might associate with treatment-resistant schizophrenia : (1) Cell-cell signaling (FDR q-value=0.0127); (2) Response to nitrogen compound (FDR q-value=0.0127); (3) Catalytic complex (FDR q-value=0.0127); (4) Reactome notch-HLH transcription pathway (FDR q-value=0.0266). **Conclusion** In our current analysis, the results showed DNA methylation differences between treatment-resistant schizophrenia and nontreatment-resistant schizophrenia patients, and that further converge into key signal pathways include cell-cell signaling, response to nitrogen compound, catalytic complex and reactome notch-HLH transcription pathway. Further analysis may help discover the roles of candidate genes linked to treatment-resistant schizophrenia.

PrgmNr 2114 - High genetic load of the *GREB1L* gene and auditory rehabilitation in the severe form of cochlear malformation

View session detail

Author Block: **B. Kim**[1], S-Y. Lee[2], J. Han[3], N. Yi[1], M. Kim[3], K. Park[3], Y. Kim[3], Y. Kim[3], B. Choi[3]; [1]Munhwa-ro 282, Chungnam Natl. Univ. Coll. of Med., Daejeon, Korea, Republic of, [2]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, [3]Seoul Natl. Univ. Bundang Hosp., Seongnam, Korea, Republic of
Disclosure Block: **B. Kim:** None.
Inner ear malformations (IEMs) are challenging in both etiologic diagnosis and treatment. Most patients with IEM have profound hearing loss, which requires cochlear implantation (CI). However, audiologic outcomes may vary. Etiologic diagnosis was elusive until recently, however recessive *ROR1* and de novo *GREB1L* variants have just been being elucidated, providing a hint of the molecular basis of this cochlear anomaly. Fueled by recent discoveries, we investigated the enigmatic genetic etiologies of such IEMs in Korea and their underlying mechanism as well as reporting the successful audiologic outcomes after CI. Four patients with IEMs were recruited. None of their parents had hearing loss. Exome sequencing was performed. CI was done in all patients and audiological outcomes were reviewed. First-line exome sequencing identified no convincing candidate variants accounting for such IEMs, however, through further bioinformatics analysis, we came across an observation that *GREB1L* variants (c.5618T>C, c.982C>T, c.1079T>A) were detected from three of four patients, which *per se* were most likely to be pathogenic and thus would have been otherwise considered a causative variant based on the population database and several in *silico* studies, were disqualified solely due to their segregation from normal-hearing mother to hearing-impaired son. Given the previous proposal of genetic imprinting for parent-origin kidney phenotype expression of this gene and also the fact that all three male probands consistently inherited the maternal mutant *GREB1L* allele in our study, variable expressivity of the hearing loss more likely suggests incomplete penetrance or genetic imprinting, rather than disqualification of these variants from candidate ones. If this is the case, the genetic load of *GREB1L* variants in this disease entity reached upto 75% in this study. Audiological outcomes showed successful auditory rehabilitation after CI. Based on the results from this study, we suggest potential high genetic load of *GREB1L* in severe IEMs and also that the parent-origin phenotype expression that potentially suggests genetic imprinting could have been the culprit for which the causative gene has not been identified for a long time in the majority of these IEMs. If this is the case, our study will be the first case series that strongly suggest genetic imprinting in this deafness field where Mendelian inheritance predominates by reporting three probands similarly exhibiting the maternal- origin expression. Contrary to concerns of outcome of CI in these IEMs, their audiological outcomes after CI are relatively good enough to enable successful language development.

PrgmNr 2115 - Mapping the cis-regulatory landscape of ABCA4 in adult human retina

View session detail

Author Block: **V. Lopez Soriano**[1], S. Van de Sompele[1], A. DueÃ±as Rey[1], E. D'haene[1], M. Bauwens[2], E. De Baere[3]; [1]Ghent Univ., Ghent, Belgium, [2]Ghent, Belgium, [3]Ghent Univ. and Ghent Univ. Hosp., Ghent, Belgium
Disclosure Block: **V. Lopez Soriano:** None.
Purpose: Stargardt disease (STGD1) is a frequent inherited retinal disease (IRD) affecting ~1/8,000 people. Significant advances have been made over the recent years in elucidating the molecular basis of STGD1, with over 600 pathogenic coding variants and with a substantial number of deep-intronic splicing variants in the disease gene ABCA4. The cis-regulatory domain of ABCA4 is unexplored so far and may represent an attractive target for non-coding disease-causing or modifying variants. By mapping and functionally validating putative cis-regulatory elements (CREs) of ABCA4, we aimed to gain more insights into the regulation of ABCA4 expression.Methods: In order to identify candidate CREs, we integrated published and in-house human retinal epigenomics datasets. CRE predictions were based on chromatin accessibility (ATAC-seq), chromosome conformation capture combined with ChIP-seq (Hi-ChIP), histone modifications (ChIP-seq) and transcriptomics data (RNA-seq), all generated on human donor retina. To functionally validate in silico predicted CREs, dual luciferase reporter assays using pGL4.23 vectors were performed in hTERT RPE-1 cells. A chromatin interaction profile of the ABCA4 locus was obtained via UMI-4C experiments on neural retina and RPE from human donors.Results: A total of 21 predicted CREs were cloned both in their native and reverse orientation in order to assess their regulatory effect in vitro. Five regions showed an increase of reporter activity, three of which display active enhancer marks (H3K4me1 and H3K27ac) in photorreceptors, and three regions showed a significant decrease in luciferase activity. The UMI-4C data showed a decrease in background compared to previously generated 4C-seq data and an improved sensitivity and resolution. The generation of replicates and reverse experiments to confirm the interactions are currently ongoing.Conclusions: Using an integrated approach based on data mining of retinal datasets, in vitro functional validation of putative retinal CREs and targeted chromosome conformation capture (UMI-4C), we have gained insight into the cis-regulatory landscape of ABCA4. The CREs identified and validated here can represent targets of non-coding pathogenic and modifying variants in cases with unsolved ABCA4-associated disease. An improved annotation of tissue-specific cis-regulatory domains of IRD genes may advance the interpretation of non-coding variants.Funding: EU H2020 ITN No. 813490.

PrgmNr 2116 - Methylation quantitative trait loci of type 2 diabetes in a middle eastern population

View session detail

Author Block: **N. Yousri**[1], S. C. Hunt[1], O. Albagha[2]; [1]Weill Cornell Med.-Qatar, Doha, Qatar, [2]Hamad Bin Khalifa Univ., Doha, Qatar
Disclosure Block: **N. Yousri:** None.
Background: With more than 15% prevalence of T2D in Qatar, it is essential to study the methylation patterns associated with T2D along with the genetic and lifestyle/environmental factors that affect methylation. GWAS studies have identified hundreds of T2D loci, yet their functional association with other biological entities still needs several years of investigation. DNA Methylation is of particular interest in this area as it has not been touched frequently when studying causality of SNPs with methylation levels. Most studies in this direction have focused on genotype arrays and earlier DNA methylation technologies. **Objectives**: We present a study on the association of DNA methylation sites with T2D in Qataris and cross that with their whole genome sequence data to investigate possible causal signals. We aim to identify CpGs associated with T2D in Qatari samples with 40% patients and identify their quantitative trait loci using whole genome data. **Materials and Methods**: 835 samples from the Qatar Bio-Bank (QBB) were profiled for >850,000 methylation sites using illumina EPIC arrays. Methylation data was normalized using 'funnorm' functions in R minfi package. CpG sites that overlap with SNPs were removed. A total of 835,000 CpG sites remained for the analysis after quality control steps. Whole genome data was filtered for a maf of >0.05, a call rate of > 98% and a hwe p value of 10^{-6}. A linear regression model for association of CpG sites with T2D was used after correction for all relevant covariates including actual cell counts that were available at the QBB, and genomic principal components. The GenABEL package was used for the CpG-SNP association analysis, correcting for kinship. Annotation data available from minfi package and Seattle-Seq server were used for the annotation. **Results**: A total of 74 CpG sites were identified to be associated with T2D at an epigenetic wide association threshold of p -8 in the discovery cohort of 454 individuals and replicated in 381 individuals. Preliminary analysis identified methylation quantitative trait loci (methQTLs) of 22 CpG sites in 29 unique genes at p -10. Half of the 22 CpG sites with methQTLs were located in the body of the CpG gene. Of all the identified methQTLs, two were *trans* associations, with the CpG site associated with SNPs in two different chromosomes. Two of the CpG sites were associated with missense and stop-gained variants, where the CpGs were located in a TSS and in exon body or gene body. More investigations are required to study and replicate the methQTLs. **Acknowledgment**: Work is supported by Qatar National Research Funds (QNRF) PPM2 #: 0226-170020, Qatar Genome Project (QGP) and Qatar Bio-Bank (QBB).

PrgmNr 2117 - Mitochondrial-nuclear gene expression co-regulation is compromised in immune system cells of COVID-19 patients: rewiring towards glycolysis

View session detail

Author Block: **H. Medini**[1], Z. Amit[1], D. Mishmar[2]; [1]Ben Gurion Univ. of the Negev, Beer Sheva, Israel, [2]Ben Gurion Univ., Beer-Sheva, Israel
Disclosure Block: **H. Medini:** None.
Mitochondria is pivotal to cellular bioenergetics and in response to viral infections. Although several SARS-CoV-2 proteins localize and interact with the mitochondria, the role of mitochondrial activities in the etiology of COVID-19 has been largely overlooked. As a first step to address this caveat, we analyzed three publicly available bulk RNA-seq datasets (two from peripheral blood and one from the respiratory tract) from COVID-19 patients and corresponding healthy controls. This analysis consistently revealed significantly reduced mitochondrial DNA (mtDNA) gene expression in peripheral blood of COVID-19 patients. In contrast, mtDNA gene expression was not significantly altered in patients' respiratory tract samples. As these findings implied cell type-dependent mitochondrial response in COVID-19 patients, we analyzed 362,758 cells from single-cells RNA-seq datasets from peripheral blood mononuclear cells (PBMC), nasopharyngeal (NP) or Broncho alveolar lavage fluid (BALF). Although significantly reduced mtDNA gene expression was observed in multiple patients' cell types, such change was more prominent in immune system cells. Notably, the reduced mtDNA gene expression associated with altered expression of candidate mtDNA transcriptional regulators, suggesting downregulation. Conversely, expression of nuclear DNA-encoded OXPHOS subunits was generally elevated in patients, suggesting a departure from mito-nuclear co-regulation. Elevated expression of ROS-response genes in patients support compromised mitochondrial function. Elevated expression of glycolysis enzymes in patients suggests rewiring towards glycolysis in patient' cells. Since SARS-CoV-2 replication requires glycolytic environment which promotes cytokine storm in immune system cells, we argue that such environment is achieved via compromised mitochondrial gene expression regulation, suggesting that mitochondrial dysfunction is central in COVID-19 etiology.

PrgmNr 2118 - UMI-4C chromatin interaction profiling, *in vitro* and *in vivo* enhancer assays to dissect the *cis*-regulatory mechanisms underlying North Carolina macular dystrophy, a retinal enhanceropathy

View session detail

Author Block: **S. Van de Sompele**[1,2], E. D'haene[1,2], B. Munevver Cicekdal[2,3], S. Vergult[1,2], T. Van der Snickt[1,2], F. S. Shaya[4], K. Vleminckx[2,3], P. Liskova[5], K. W. Small[4], E. De Baere[1,2]; [1]Dept. of Biomolecular Med., Ghent Univ., Ghent, Belgium, [2]Ctr. for Med. Genetics, Ghent Univ. Hosp., Ghent, Belgium, [3]Dept. of BioMed. Molecular Biology, Ghent Univ., Ghent, Belgium, [4]Macula and Retina Inst., Los Angeles and Glendale, CA, [5]Res. Unit for Rare Diseases, Charles Univ. and Gen. Univ. Hosp., Prague, Czech Republic

Disclosure Block: **S. Van de Sompele:** None.

North Carolina macular dystrophy (NCMD) is a rare autosomal dominant retinopathy, characterized by loss of central vision. With the identification of noncoding single nucleotide variants (SNVs) and duplications overlapping with a DNase I hypersensitive site (DHS) near *PRDM13* or *IRX1* as the genetic cause of NCMD, it is hypothesized to be a retinal enhanceropathy. Here we aim to dissect the *cis*-regulatory domains of *PRDM13* and *IRX1*, to provide mechanistic insight into the molecular pathogenesis of NCMD.

Chromosome conformation capture, in particular UMI-4C sequencing, was performed on retinas from human donor eyes to fine-map interactions of *cis*-regulatory elements (CREs) with the *PRDM13* and *IRX1* promoters. This data was integrated with other human retinal (epigen)omics data (ChIP-seq of histon modifications and retinal transcription factors, ATAC-seq, RNA-seq, DNase-seq of embryonic retina at five different stages) in a genome-wide UCSC browser session to identify candidate retinal CREs. Twenty-five unrelated NCMD families underwent targeted testing followed by whole genome sequencing (WGS) in the unsolved cases (n=11). Subsequently, *in vitro* luciferase assays in ARPE-19 cells and *in vivo* enhancer detection assays in Xenopus leavis using a ZED vector were performed to evaluate candidate CREs and (likely) pathogenic SNVs.

This multi-omics analysis mapped the retinal chromatin interaction profile of both disease loci, and revealed 7 and 6 candidate CREs for *PRDM13* and *IRX1* respectively, two of which are hotspots of pathogenic SNVs. The genetic architecture of NCMD was expanded with a novel noncoding SNV and duplication with a likely effect on *PRDM13* regulation. Interestingly, UMI-4C showed an interaction of the CRE this variant is located in with the PRDM13 promoter. This CRE is active at a specific developmental stage (D103) compatible with the timepoint when retinal progenitor cells of the central retina exit mitosis. Finally, luciferase assays demonstrated that the known and novel noncoding SNVs, located in the two hotspots, display higher activation and silencing, respectively. Overall, we have provided insight into the *cis*-regulatory mechanisms underlying NCMD, an examplary retinal enhanceropathy.

PrgmNr 2119 - ZNF143, an important factor for CTCF-bound chromatin interactions

View session detail

Author Block: **Q. Zhou**[1], M. Yu[2]; [1]Natl. Univ. of Singapore, Singapore, Singapore, [2]Univ. of California San Diego
Disclosure Block: **Q. Zhou:** None.
CCCTC binding factor (CTCF) is an important factor in the maintenance of chromatin-chromatin interactions, yet the mechanism regulating its binding to chromatin is unknown. We demonstrate that zinc finger protein 143 (ZNF143) is a key regulator for CTCF-bound promoter-enhancer loops. In the murine genome, 10,544 pairs of ZNF143 and CTCF motifs located 37 bp away from each other, and 99.9% (10,529) formed the convergent orientation. Furthermore, deletion of ZNF143 leads to loss of CTCF binding on promoter and enhancer regions associated with gene expression changes. CTCF-bound promoter-enhancer loops are also disrupted after excision of ZNF143. Furthermore, deletion of ZNF143 in hematopoietic system leads to embryonic lethal, suggesting an essential role in hematopoiesis. Indeed, systematically analysis demonstrates that both embryonic and adult hematopoietic stem cells completely loss the ability to reconstitute the hematopoietic system in recipient mice without ZNF143. In line with that multiple pathways and genes critical for hematopoiesis are downregulated, associated with significant reduction of CTCF binding on their promoter-enhancer loop anchors and decreased loop intensities, we hypothesize that CTCF-bound promoter-enhancer loops regulate gene expression patterns essential for maintenance of murine hematopoietic stem and progenitor cell integrity. Our data suggest a common feature of gene regulation that ZNF143 is a critical factor for CTCF-bound promoter-enhancer loops.

PrgmNr 2120 - Epigenome-wide association study of the human IgG N-glycome composition

View session detail

Author Block: **A. Frkatovic**[1], F. Vuckovic[1], J. F. Wilson[2,3], G. Lauc[1,4], L. Klaric[2]; [1]Genos GlycoSci. Res. Lab., Zagreb, Croatia, [2]MRC Human Genetics Unit, Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom, [3]Usher Inst., Univ. of Edinburgh, Edinburgh, United Kingdom, [4]Faculty of Pharmacy and Biochemistry, Univ. of Zagreb, Zagreb, Croatia
Disclosure Block: **A. Frkatovic:** None.
N-glycosylation of immunoglobulin G (IgG) is a post-translational modification which affects protein stability and its effector functions. IgG N-glycosylation is influenced by both genetic and environmental factors. Epigenetic mechanisms, such as DNA methylation, may provide further insight into the link between IgG N-glycome heritability and environment. We performed an epigenome-wide association study (EWAS) in the ORCADES cohort of volunteers from Northern Scotland (n=967) to identify CpG sites associated with the composition of the IgG N-glycome, quantified by ultra-performance liquid chromatography. Nine glycan traits were defined to reflect biologically relevant traits by summing up the glycan structures containing a specific sugar unit in the total IgG N-glycome, such as fucose. Genome-wide methylation levels were measured from whole blood DNA samples with the Illumina HumanMethylation 850K BeadChip, followed by QC and correction for technical and biological covariates. Linear regression was performed with each glycan trait as outcome and CpG M value as the independent variable, with correction for age and sex in the basic model. The adjusted model further included smoking, BMI, and age-by-sex interaction. In addition, we performed a sex-stratified test. The significantly associated CpG sites were assessed for potential confounding by the surrounding SNPs (Â±50 kb). At p-8, basic and adjusted models resulted in eight and five associated CpG sites, respectively. The association of cg08178031 and sialylation ($p=4.39x10^{-9}$) was confounded by a SNP in the 5â□□UTR region of the *ST6GAL1* gene which is known for its role in the addition of sialic acid to the glycan chain. In a sex-stratified test, we demonstrated that the association of galactosylation and cg06072257 is sex-specific (female-only), noting that this association is only suggestive ($p=3.17x10^{-7}$), likely due to lower power for the smaller sample size in the sex-stratified analysis. Results of eFORGE tool for exploration of transcription factor binding sites show that cg06072257 is found near to the binding motif for ESR1. Two additional associations were significant in the female-specific test: cg08946781 (with sialylation and galactosylation) found in the open sea region next to the *CCR7* gene and cg27419075 (bisection) located in a CpG island upstream of the *PDE2A* gene. A single male-specific association (cg16564136) mapped to the 5â□□UTR region of the *NMNAT3* gene. Despite the limited power, there is an implication that the methylation of the associated CpG sites might be sex-specific, thereby affecting the composition of the IgG N-glycome in a sex-specific manner.

PrgmNr 2121 - eQTL Catalogue: A compendium of uniformly processed human gene expression and splicing QTLs

View session detail

Author Block: **N. Kerimov**[1,2], J. D. Hajhurst[3,2], K. Peikova[1], J. R. Manning[3,2], P. Walter[3], L. Kolberg[1], M. Samovica[1], M. P. Sakthivel[3,2], I. Kuzmin[1], S. J. Trevanion[3,2], T. Burdett[3,2], S. Jupp[3,2], H. Parkinson[3,2], I. Papatheodorou[3,2], A. D. Yates[3,2], D. R. Zerbino[3,2], K. Alasoo[4,2]; [1]Inst. of Computer Sci., Univ. of Tartu, Tartu, Estonia, [2]Open Targets, Hinxton, Cambridge, United Kingdom, [3]EMBL-EBI, Hinxton, Cambridgeshire, United Kingdom, [4]Univ. of Tartu, Tartu, Estonia

Disclosure Block: **N. Kerimov:** None.

An increasing number of gene expression quantitative trait locus (eQTL) studies have made summary statistics publicly available, which can be used to gain insight into complex human traits by downstream analyses, such as fine mapping and colocalisation. However, differences between these datasets, in their variants tested, allele codings, and in the transcriptional features quantified, are a barrier to their widespread use. Consequently, target genes for most genome-wide association study (GWAS) signals have still not been identified. Here, we present the eQTL Catalogue (https://www.ebi.ac.uk/eqtl/), a resource that contains quality controlled, uniformly re-computed QTLs from 21 eQTL studies. We find that for matching cell types and tissues, the eQTL effect sizes are highly reproducible between studies, enabling the integrative analysis of these data. Although most cis-eQTLs were shared between most bulk tissues, the analysis of purified cell types identified a greater diversity of cell-type-specific eQTLs. Consequently, the number of independent GWAS loci for which we could detect at least one colocalising eQTL increased by 25% when we added eQTL Catalogue datasets to the GTEx v8. Our summary statistics can be downloaded by FTP, accessed via a REST API, and visualised on the Ensembl genome browser. Fine mapped associations (credible sets) can also be explored using our user-friendly website (https://elixir.ut.ee/eqtl/). New datasets will continuously be added to the eQTL Catalogue, enabling the systematic interpretation of human GWAS associations across many cell types and tissues.

PrgmNr 2122 - Evaluation of the impact of genetic diversity on PIWI-interacting RNA analysis

View session detail

Author Block: **K. Sato**[1,2], K. Ide[3,4,5], N. Tanaka[5]; [1]Systems Aging Sci. and Med., Tokyo Metropolitan Inst. of Gerontology, Tokyo, Japan, [2]Integrated Res. Initiative for Living Well with Dementia, Tokyo Metropolitan Inst. of Gerontology, Tokyo, Japan, [3]Dept. of Life Sci. and Med. BioSci., Waseda Univ., Tokyo, Japan, [4]Computational Bio Big-Data Open Innovation Lab. (CBBD-OIL), Natl. Inst. of Advanced Industrial Sci. and Technology, Tokyo, Japan, [5]Data Sci. Res. Section, Hlth.y Aging Innovation Ctr., Tokyo Metropolitan Inst. of Gerontology, Tokyo, Japan
Disclosure Block: **K. Sato:** None.
Approximately 45% of the human genome are occupied by transposable elements or transposons, which are selfish mobile DNA sequences, translocating around the genome. The translocation of transposons, also refer as transposition often leads severe genic breakage and chromosomal instability, consequently the species conservation is threatened. In the germline cells, specialized machineries to repress transposons have been acquired, which is PIWI-interacting RNA (piRNA)-mediated RNA silencing. piRNA is a largest class of small non-coding RNA and is mostly arisen from the specified genomic loci, called piRNA clusters, which usually harbors transposon fragments and remnants, thereby targeting transposons. Although piRNAs are predominantly expressed in the germline, the elevated expression has also been observed in somatic cells and cancer cells in animal. Transposons, in particular retrotransposons, move around the genome by copy and paste mechanism, thereby amplifying their own sequences in number quickly. Importantly, on the piRNA analysis, the transposition causes genetic diversity between individuals, which may interrupt the bioinformatics analyses such as reduced mappability, incompatibility between samples, and missing orphan piRNAs. In order to assess the impact of genetic diversity on piRNA analysis, we first sought small RNA-seq data sets from normal human testis and found three data sets from the testis of Caucasians and a data set isolated in China. We next defined 697 common piRNA clusters between two human reference genomes, standard reference genome GRCh38/hg38 and Han Chinese reference genome HG00514, and then performed the mapping of the data sets to these reference genomes. Note that GRCh38/hg38 is mostly based on Caucasian and African ancestry. Unexpectedly, all four data sets were mapped to HG00514 higher than GRCh38/hg38, suggesting that although the number of samples is small in this study, the nucleotide diversity may not be an important factor in determining the mappability of piRNA reads, and further suggesting that piRNA reads would be better to be analyzed utilizing several reference genomes to avoid the missing of functional piRNAs in individuals.

PrgmNr 2123 - Loss of PTEN disrupts 5-hydroxymethylcytosine landscape along mESC neuronal differentiation

View session detail

Author Block: **K. Miu**, Z. Wang, S. Fung, C. Yu, C. Huang, W. Wang, F. Ou, S. Hao, W. Law, W-Y. Chan; The Chinese Univ. of Hong Kong, Hong Kong SAR, China
Disclosure Block: **K. Miu:** None.
DNA hydroxymethylation dynamics have been implicated in normal development and differentiation. 5-hydromethylcytosine (5hmC), created by Tet proteins-catalysed oxidation of 5-methylcytosine (5mC), is most abundantly found in the brain, but the genome-wide 5hmC landscape along diverse neuronal differentiation remains unknown. Interestingly, we found that Pten ablation effectively disrupted the lineage differentiation of midbrain/hindbrain neural progenitors, resulting in immature neuronal progenitors.Here, we differentiated mouse embryonic stem cells (mESCs) into ventral midbrain and hindbrain neural progenitors and characterized the global 5hmC distribution using nano-5hmC-seal approach. The 5hmC pattern was found remodelled in promoters, exons and enhancers, associated with gene activation and repression. Notably, ventral midbrain marker (Lmx1a, Otx1, Th) and hindbrain marker (Hoxa1, Zic1, Tph1) acquire 5hmC and thereby up-regulated during differentiation. Furthermore, dynamic changes in 5hmC within the same neural lineages were investigated when the mESCs were subjected to Pten ablation. Indeed, we identified 4111 and 203 differentially hydroxymethylated regions (DhMRs) in differentiation of Pten-/- mESCs into ventral midbrain and hindbrain progenitors. Gene ontology analysis showed that the majority of these DhMRs were associated with neurogenesis and relevant signal transduction pathways along ectoderm differentiation. To conclude, our findings demonstrate active hydroxymethylation remodels lineage-specific differentiation of pluripotent stem cells during early neural development.

PrgmNr 2124 - Nuclease deficiencies alter plasma cell-free DNA methylation profiles

View session detail

Author Block: **D. Han**[1], M. Ni[1], R. W. Y. Chan[1], D. K. L. Wong[1], L. T. Hiraki[2], S. Volpi[3], P. Jiang[1], K. O. Lui[1], K. C. A. Chan[1], R. W. K. Chiu[1], Y. M. D. Lo[1]; [1]The Chinese Univ. of Hong Kong, Shatin, Hong Kong, [2]The Hosp. for Sick Children, Toronto, ON, Canada, [3]Istituto Giannina Gaslini, Genova, Italy
Disclosure Block: **D. Han:** None.
The effects of DNASE1L3 or DNASE1 deficiency on cell-free DNA (cfDNA) methylation was explored in plasma of mice deficient in these nucleases and in DNASE1L3-deficient humans. Compared to wildtype cfDNA, cfDNA in *Dnase1l3*-deficient mice was significantly hypomethylated, while cfDNA in *Dnase1*-deficient mice was hypermethylated. The cfDNA hypomethylation in *Dnase1l3*-deficient mice was due to increased fragmentation and representation from open chromatin regions (OCRs) and CpG islands (CGIs). These findings were absent in *Dnase1*-deficient mice, demonstrating the preference of DNASE1 to cleave in hypomethylated OCRs and CGIs. We also observed a substantial decrease of fragment ends and coverage at methylated CpGs in the absence of DNASE1L3, thereby demonstrating that DNASE1L3 prefers to cleave at methylated CpGs. Furthermore, we found that methylation levels of cfDNA varied by fragment size in a periodic pattern, with cfDNA of specific sizes being more hypomethylated and enriched for OCRs and CGIs. These findings were confirmed in DNASE1L3-deficient human cfDNA. Thus, we have found that nuclease-mediated cfDNA fragmentation markedly affected cfDNA methylation level on a genome-wide scale. This work provides a foundational understanding of the relationship between methylation, nuclease biology and cfDNA fragmentation.

PrgmNr 2125 - Regulon active landscape reveals cell development and function state changes of human primary osteoblasts *in vivo*

View session detail

Author Block: **S. Wang**[1], Y. Gong[2], Z. Wang[3], X. Li[4], X. Qiu[1], H. Zhang[4], Y. Liu[4], L. Cheng[5], Y. Hu[6], L. Tan[4], X. Meng[1], H. Shen[2], H. Xiao[1], H. Deng[2]; [1]Ctr. of Reproductive Hlth., System Biology and Data Information, Inst. of Reproductive & Stem Cell Engineering, Sch. of Basic Med. Sci., Central South Univ., Changsha, China, [2]Tulane Ctr. of BioMed. Informatics and Genomics, Deming Dept. of Med., Tulane Univ. Sch. of Med., Tulane Univ., New Orleans, LA, [3]Xiangya Nursing Sch., Central South Univ., Changsha, China, [4]Lab. of Molecular and Statistical Genetics, Coll. of Life Sci., Hunan Normal Univ., Changsha, China, [5]Dept. of Orthopaedics and Natl. Clinical Res. Ctr. for Geriatric Disorders, Xiangya Hosp., Central South Univ., Changsha, China, [6]Dept. of Orthopedics, Xiangya Hosp., Central South Univ., Changsha, China

Disclosure Block: **S. Wang:** None.

Transcription factor (TF) regulation played an important role in human osteoblastsâ function. The molecular features of TFs in human osteoblasts were not fully understood. We used single-cell RNA sequencing (scRNA-seq) to perform a systematic cellular taxonomy dissection of freshly isolated human osteoblasts. Then, based on human hg19 transcriptional regulator database, we calculated regulation modules by single-cell regulatory network inference and clustering (SCENIC). Additionally, we performed cell-specific network (CSN) analysis among 4 important regulons (*CREM, FOSL2, RUNX2* and *CREB3L1*), and the total number of significant edges in the whole transcriptome CSN was returned as the network degree matrix (NDM) value for each TF. Then, we constructed regulon activity-based osteoblast development trajectories by Monocle (v2.14.0). Our research revealed cell development and function state changes of primary osteoblasts by using SCENIC analysis and CSN construction. *CREM* and *FOSL2* regulons were relatively active in preosteoblast-S1. Strong CSN connections of immunity, cell proliferation/differentiation related target genes in *CREM*, *FOSL2* regulons further proved their particular functional state. Our results also support that osteoblast differentiation and development were influenced by different TFs in different cell stage. For example, the key time point of *RUNX2* and *CREB3L1*âs regulation may be in the mature stage of osteoblast. As expected, preosteoblast-S1 showed the highest NDM value of *CREM* and *FOSL2*, mature osteoblast showed the highest NDM value of *RUNX2* and *CREB3L1,* which indicated their wide influence at the whole transcriptome level through directly regulating gene targets or other indirect interactions, co-expression, alternative splicing and so on. This is the first study to describe the unique features in cellular regulon active landscape of human osteoblasts *in vivo* and identify the gene associations/network for osteogenesis at single-cell resolution. Our results may provide valuable resources to reveal the advanced mechanisms of bone metabolism and associated disorders based on the TF regulation network.

PrgmNr 2126 - TDG contributes to male subfertility through dysregulation of male germline stem cell development

View session detail

Author Block: **T-L. Lee**, J. Liao, W. Lee, H. Suen, K. Chow, A. Luk; The Chinese Univ. of Hong Kong, Shatin, China
Disclosure Block: **T. Lee:** None.
Thymine DNA glycosylase (TDG) is a key enzyme in active DNA demethylation, which could excise 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). To investigate the impact of *Tdg* on male fertility, we created a mouse strain with male germline-specific *Tdg* knockout using Stra8-Cre system. Our breeding assay showed *Tdg*-conditional knockout (*Tdg*-CKO) males had fewer accumulated progenies over a 35-week period compared to heterozygous males, indicating sub-fertility. Testis section showed complete spermatogenesis in *Tdg*-CKO males but an abnormally lower number of spermatocytes in the leptotene/zygotene stages. TUNEL assay also showed an increased number of apoptotic cells in *Tdg*-CKO testes. Meanwhile, our RNA-Seq analysis on neonatal KIT- spermatogonial stem cells (SSC) identified 110 differentially expressed genes (DEGs) between KO and WT. Interestingly, *Tdg* deletion led to suppression of genes in the FGF signaling pathway while enhancement in the competing GDNF pathway in undifferentiated SSC. Finally, ATAC-Me showed hypo-demethylation during SSC differentiation (KIT transition) in promoter regions of *Id4* and *Solhl1*, which are important spermatogenesis regulators, suggesting *Tdg* is regulating site specific demethylation.

PrgmNr 2127 - Whole-genome and small RNA sequencing-based microRNA-eQTL mapping in Japanese elucidates variant-microRNA-disease connections

View session detail

Author Block: **K. Sonehara**[1], S. Sakaue[1], Y. Maeda[1], J. Hirata[1], T. Kishikawa[1], K. Yamamoto[1], H. Matsuoka[2], M. Yoshimura[2], T. Nii[1], S. Ohshima[2], A. Kumanogoh[1], Y. Okada[1]; [1]Osaka Univ. Graduate Sch. of Med., Suita, Japan, [2]NHO Osaka Minami Med. Ctr., Kawachinagano, Japan
Disclosure Block: **K. Sonehara:** None.
Expression quantitative trait loci (eQTL) analysis of non-coding RNA (ncRNA) is an essential step toward elucidating the ncRNA involvement behind the genetics of complex diseases. Among several classes of ncRNAs, microRNAs (miRNAs) are key post-transcriptional gene regulators and implicated in diverse diseases. Despite its biological importance, existing miRNA-eQTL studies have focused mostly on European individuals, underrepresented in other populations. As an initial miRNA-eQTL study in the Asian population, we conducted miRNA-eQTL mapping for 343 miRNAs in 141 Japanese individuals using small RNA sequencing (sRNA-seq) of peripheral blood mononuclear cells (PBMCs) and whole-genome sequencing (WGS). We identified 1,275 *cis*-miRNA-eQTL variants for 40 miRNAs (false discovery rate cis-regulatory variants not captured by the conventional miRNA-eQTL mapping (e.g., miR-933). We identified a copy number variation (CNV) associated with miRNA expression (e.g., miR-570-3p, P = 7.2 Ã 10^{â6}), which contributes to a more comprehensive landscape of miRNA-eQTLs. To elucidate a post-transcriptional modification in miRNAs, we created a catalog of miRNA-editing sites, including ten canonical (A-to-I and C-to-U) and six non-canonical sites (other base substitutions). Finally, by integrating the miRNA-eQTLs and the large-scale Japanese genome-wide association studies (GWASs) of 25 complex traits (mean n = 192,833), we conducted a transcriptome-wide association study (TWAS) and colocalization analysis. We identified miR-1908-5p as a potential mediator for adult height, colorectal cancer, and type 2 diabetes (P â5). Our study broadens the population diversity in ncRNA-eQTL studies and contributes to functional annotation of disease-associated loci found in non-European populations.

PrgmNr 2128 - Identification of hubb genes in retinoblastoma using network analysis

View session detail

Author Block: **S. Bhatia**; Univ. of South Florida, Tampa, FL
Disclosure Block: **S. Bhatia:** None.
Background: Retinoblastoma (RB) is a rare cancer usually initiated by biallelic mutation of the retinoblastoma gene (*RB1*). This study was performed to identify genes related to RB using network-based analysis. **Methods:** The differentially expressed genes (DEGs) were analyzed using GEO2R from GSE97508 of the Gene Expression Omnibus (GEO) database. Gene enrichment analysis, and protein-protein interaction (PPI) network construction were performed on the DEGs by the Gene Ontology (GO), Search tool for the retrieval of interacting Genes (STRING), and Cytoscape to identify hub genes. Finally, the network was analyzed and genes were ranked based on several centrality indices. **Results:** A total of 250 DEGs were identified that mainly play role in visual transduction processes. The networks were analyzed and high centrality genes were identified. Exploration of functional modules and complexes showed that the majority of high centrality genes integrated in biological pathways that drive retinoblastoma pathogenesis. **Conclusions:** In this study, the network-based analysis was proposed that could explicate the complex interplay between biological processes underlying RB. Furthermore, an experimental validation of these candidate genes could lead to identification of promising targets.

PrgmNr 2129 - Detecting gene-gene interactions from GWAS using diffusion kernel principal components

View session detail

Author Block: **A. Walakira**[1], J. Ocira[2], D. Duroux[2], R. Fouladi[2], M. MoÅikon[1], D. Rozman[1], K. Van Steen[2]; [1]Univ. of Ljubljana, Ljubljana, Slovenia, [2]Univ. of Liege, Liege, Belgium
Disclosure Block: **A. Walakira:** None.
Genes and gene products function as components of complex networks of macromolecules through physical or biochemical interactions. Dependencies of gene mutations on genetic background (i.e., epistasis) play a role in understanding molecular underpinnings of complex diseases like inflammatory bowel disease (IBD). However, the process of identifying such interactions is complex due to the curse of high dimensionality, dependencies in the data and non-linearity. We propose a novel epistasis detection analysis workflow that 1) takes GWAS SNP data as input, 2) develops gene-level summaries via diffusion kernels on graphs, and 3) uses a Bayesian framework and kernel PCs as new units of analysis to discover gene-gene interactions in relation to a complex trait.We applied our analysis workflow on GWAS data on IBD from the IIBDGC consortium (www.ibdgenetics.org). We pruned out SNPs that were in linkage disequillibrium ($r^2 > 0.75$) and considered only common variants (minor allele frequency > 5%) and those in Hardy-Weinberg equilibrium (p-value > 0.001). We used FUMA to functionally map SNPs to genes (significant eQTL SNP-gene expression in colon). Bivariate synergy was used to construct SNP-based graphs within a gene and the resulting within-gene edge weights were used to obtain informed summaries of genes. For analysis of within-gene network, the 'Maximum Relevance Minimum Redundancy' algorithm in the minet package in R was used for meaningful node selection. This resulted in a reduced gene SNP set that was analysed with the igraph R package. Furthermore, for each gene, diffusion kernels were constructed over genotypes based on matrix exponentiation. The first principal component for each gene was retained and considered a â□□summaryâ□□ explanatory variate in Bayesian gene-gene interaction models. Interactions were highlighted based on posterior inclusion probabilities greater than zero.Generally, within gene networks differed among genes, with genes such as *PARK7, CD40, NUCKS1, MST1R* and *SLC22A4,* known to be associated with IBD (CD or UC), showing higher-end densities. Gene size (number of SNPs) was not the determining factor for epistasis detection. Seven genes involved in 2-way interactions namely *LINCO1475, TAP2, RGS14, OTUD3, SLC22A4, NICN1* and *NOTCH4* have been associated with IBD, CD or UC in DAVID and DisGeNET. Our method identified four genes namely MAP1LC3A, RGS14, CNFTR and AHSA2 to be of interest in IBD. We show that our approach is able to recover known IBD associated genes and additional genes of interest linked to disease. The trait-informed dimensionality reduction step in our analysis workflow enhances the detection of gene-gene interaction effects.

PrgmNr 2130 - Epigenetic and RNA profiling of cancer using multi-omics data on the Cancer Genomics Cloud powered by Seven Bridges

View session detail

Author Block: **V. Pajic**[1], N. Vukojicic[1], J. Babic[1], A. Danicic[1], M. Matic[1], A. Stankovic[1], D. A. Dean,II[2], J. W. Grover[2], M. Ray[2], S. Subramanian[3]; [1]Seven Bridges, Belgrade, Serbia, [2]Seven Bridges, Boston, MA, [3]Seven Bridges, Sugar Land, TX

Disclosure Block: **V. Pajic:** None.

Integrative analysis of multi-omics data is a promising approach for revealing the functionality of complex biological systems and processes in cancer. The Cancer Genomics Cloud (CGC) powered by Seven Bridges is an NCI-funded computational platform for accessing and analyzing large cancer datasets and facilitating collaborative research on the cloud. It is a comprehensive resource that provides access to complex datasets generated by diverse omics technologies and computational tools that can be used for their analysis. In order to enable multi-omics analysis on CGC, we present an end-to-end bioinformatics pipeline for processing epigenetic data on the CGC.The customizable pipeline is modular and implemented in the Common Workflow Language (CWL) on the CGC. A user can initiate the analysis from either raw FASTQ files coming from RNAseq, ATACseq and ChIPseq experiments, or from already processed data (gene expression tables or peak lists). If raw reads are used, the files are first processed with an adequate set of tools, specific for each type of experiment. For RNAseq, we use a pipeline that includes basic quality control, alignment and quantification steps. ATACseq and ChIPseq data are processed based upon ENCODE consortium specification, which includes reads alignment, peak calling, quality control and reporting steps.The resulting files, expression tables from RNAseq experiments and peaks from ATACseq and ChIPseq, are further jointly processed within the pipeline. This joint analysis starts with comparing the transcript count and peak set data to identify differentially expressed genes and binding regions. The detected differential features are then overlapped and used for gene set enrichment and pathway analysis, in order to explore the correlation of chromatin accessibility or transcription factor binding and gene expression in cancer samples. The analysis is implemented in two forms: as a workflow that can be used as a one-click solution through the CGC webuser interface, or as an interactive analysis through an R studio notebook, also available on the CGC.In summary, we present a bioinformatics pipeline that allows researchers to study the epigenetic mechanisms involved in transcriptional regulation in cancer. The CGC empowers the users to complete their entire research workflow on the platform while streamlining collaboration and speeding the time from hypothesis to conclusion.

PrgmNr 2131 - Implementation and evaluation of available methods for epimutation analysis

View session detail

Author Block: **C. Ruiz Arenas**[1,2], L. Abarrategui[3,4,5], C. Hernandez-Ferrer[6,7], L. A. Perez-Jurado[8,2,9], J. R. GonzÃ¡lez[3,8,10]; [1]Unitat de Genetica, DCEX, Univ.t Pompeu Fabra, Barcelona, Spain, [2]Centro de InvestigaciÃ³n BiomÃ©dica en Red de Enfermedades Raras CIBERER, Barcelona, Spain, [3]Bioinformatics Res. Group in Epidemiology (BRGE), Barcelona Inst. for Global Hlth.(ISGlobal), Barcelona, Spain, [4]Faculty of Med. Sci., Newcastle Univ., Newcastle-Upon-Tyne, United Kingdom, [5]Univ.t Autonoma de Barcelona, Bellaterra (Cerdanyola del VallÃ©s), Spain, [6]Ctr. Natl. dâ□□AnÃ lisi GenÃ²mica (CNAGâ□□CRG), Barcelona, Spain, [7]The Barcelona Inst. of Sci. and Technology, Barcelona, Spain, [8]Univ.t Pompeu Fabra, Barcelona, Spain, [9]Genetics Service, Hosp. del Mar Res. Inst., Barcelona, Spain, [10]CIBER Epidemiologa y Salud PÃºblica (CIBERESP), Barcelona, Spain

Disclosure Block: **C. Ruiz Arenas:** None.

Epimutations are rare alterations in the methylation pattern at specific loci. Epimutations can lead to cancers, such as Lynch syndrome, rare diseases such as Prader-Willi syndrome, and are associated with common disorders, such as autism. Nonetheless, no standard methods are available to detect and quantify these alterations. Two methods for epimutations detection on methylation microarray data have been reported: (1) based on identifying CpGs with outlier values and then cluster them in epimutations (Barbosa et al, 2018); (2) define candidate regions with bumphunter and test their statistical significance with a MANOVA (Aref-Eshghi et al, 2019). However, the implementation of these methods is not publicly available and these approaches have not been compared. We have implemented these two methods (called â□□barbosaâ□□ and â□□manovaâ□□, respectively) in the epimutacions R package. Additionally, we implemented four additional approaches, using a different distribution to detect CpG outliers (beta), or a different model to assess region significance (mlm, mahdistmcd, and isoforest). epimutacions can infer epimutations based on a case-control design, or using a leave-one-out approach, conforms with Bioconductor guidelines and is publicly available at https://github.com/isglobal-brge/epimutacions. First, we explored the effect of sample size on the detection of 4 epimutations experimentally validated from 2 GEO datasets. In this setting, methods based on bumphunter (manova, mlm and mahdistmcd) had a higher recall. Next, we run all methods in Illumina 450K methylation microarray data from 378 DNA samples from cord blood of the INMA cohort, a general population cohort. beta, barbosa, mlm, and manova identified few epimutations per individual, while mahdistmcd and isoforest detected tens or hundreds of epimutations per individual. Next, we assessed the effect of normalization algorithms and the inclusion of different batches on epimutation detection. In both cases, methods based on CpG outlier identification (beta and barbosa) returned more consistent results. Finally, we run beta, barbosa, mlm, and manova in the DNA methylation microarray data from blood of 210 children of INMA (4 years old) and 860 children of HELIX (8 years old), a general cohort containing 220 children from INMA. We observed a lower burden of epimutations in blood DNA of children (1-15%) than in cord blood DNA (18-30%), and we found

PrgmNr 2132 - Pathway analysis enhances characterization of cell types and sample groups in single-cell RNA sequencing

View session detail

Author Block: **T. Kakati**[1,2], Z-X. Liu[3], D. Bhattacharyya[1], J. Kalita[4], T. R. Morgan[5], T. M. Norden-Krichmar[6]; [1]Tezpur Univ., Tezpur, India, [2]Univ. of California Irvine, Irvine, CA, [3]Keck Graduate Inst., Claremont, CA, [4]Univ. of Colorado, Colorado Springs, Colorado Springs, CO, [5]VA Long Beach Hlth.care System, Long Beach, CA, [6]Univ of California Irvine, Irvine, CA
Disclosure Block: **T. Kakati:** None.
Introduction: Non-alcoholic fatty liver disease (NAFLD) and alcohol-related liver disease (AC) share similar pathologic findings although the inciting agents are different. We sought to better characterize the biological pathways involved in these two diseases using single-cell RNA sequencing (scRNA-seq) data. **Methods:** This study was conducted on peripheral blood mononuclear cells (PBMCs) from patients enrolled and consented for an IRB-approved study in the Southern California Alcoholic Hepatitis Consortium (SCAHC) with alcohol-related cirrhosis (AC, n=1), non-alcoholic fatty liver disease (NAFLD, n=1), and healthy control (HC, n=1). The PBMC samples were prepared for 10X Genomics scRNA-seq, and sequenced on an Illumina platform. Seurat version 4.0.1 was used to perform the integrated analysis. Ingenuity Pathway Analysis (IPA) was used for pathway analysis of the differentially expressed genes (DEGs) found by comparing the three comparison groups: AC vs. HC, AC vs. NAFLD, and NAFLD vs. HC. We identified common and unique pathways on the basis of most significant p-values and z-scores across the different cell types and comparison groups. **Results:** We identified 9 cellular subgroups in the PBMC samples using cell-type markers. The significant DEGs for each cell type and disease comparison group were input into IPA. We found significant pathways which were associated with immune response, inflammation, and fibrosis. Some common pathways were identified across multiple cell types, such as: Integrin signalling (B and NK cell types), regulation of Actin-based motility by Rho (Memory T, B, and NK cell types), and RhoA Signalling (CD14 monocytes, B, and NK cell types). Additionally, we found that certain unique pathways were only enriched in a particular cell type of a comparison group. For example, for AC vs. HC, IL-15 signalling was found in B cells. For AC vs. NAFLD, chemokine signalling was significant in FCGR3A monocytes, and IL-7 signalling in CD8T cell type. For NAFLD vs. HC, IL-23 signalling was significant in Memory T cells, and Interferon signalling in NK cells. These pathways play major roles in AC and NAFLD and have the potential to distinguish between diseases and cell types. **Conclusion:** Using scRNA-seq data analysis from PBMCs, we identified several pathways mapped from DEGs and cell types in two common liver diseases, AC and NAFLD. Understanding of physiological parameters and mechanism of liver diseases at the single cell level will be critical for development of new and effective treatments for alcohol-associated and non-alcohol-associated liver diseases.

PrgmNr 2133 - Spatial changes in the human genome indicated by the consensus-based structural variants in selected human families

View session detail

Author Block: **M. Chilinski**[1,2], S. Gadakh[1], P. Dariusz[1,2]; [1]Ctr. of New Technologies, Univ. of Warsaw, Warsaw, Poland, [2]Faculty of Mathematics and Information Sci., Warsaw Univ. of Technology, Warsaw, Poland

Disclosure Block: **M. Chilinski:** None.

In this work, we present a comprehensive analysis of spatial changes indicated by the genomic variability. We have used Oxford Nanopore sequencing technology along with short-read Illumina, focusing on detection of the structural variants (SVs). SVs are DNA segments of at least 50 bp in length that are often unique for personal genomes, as identified by various studies, including the 1000 Genomes project. We present an attempt to improve the quality of the Structural Variants identification from the whole genome sequencing (WGS) experiments by using the consensus approach. The biological samples being analysed originate from the 1000 Genomes Project (three healthy Trios, specifically daughters from those three families). Fifteen gold-standard silico callers were used for obtaining the polished list of Structural Variants for each family; these were obtained from next generation sequencing experiments performed by both short-read (Illumina) and long-read (Oxford Nanopore). The results of the SV callings were merged using the novel ConsensuSV algorithm, which integrates the SV sets using machine learning by combining decision trees and neural networks trained and benchmarked on the high-quality SVs from the The Human Genome Structural Variation Consortium. We provide the validated sets of high-confidence Structural Variants identified for each of the analysed daughters from Trios. Having identified the SVs from the both short-read and long-read techniques, we have mapped them onto reference (GM12878) high-order structures, including chromatin contact domains (CCDs) and ChIA-PET loops, and compared them with in-situ ChIA-PET data for the given cell lines. The case studies provide insight on the changes of spatial chromatin conformation, potentially related to the Structural Variants that are present in the given sites.

PrgmNr 2134 - CNV characteristics in Chernobyl power plant catastrophe clean-up workers from Lithuania suggest unique genetic variation structure

View session detail

Author Block: **I. Domarkiene**, G. Zukauskaite, A. Matuleviciene, A. Urnikyte, V. Kucinskas, L. Ambrozaityte; Dept. of Human and Med. Genetics, Inst. of BioMed. Sci., Faculty of Med., Vilnius Univ., Vilnius, Lithuania

Disclosure Block: **I. Domarkiene:** None.

Background. Copy number variants (CNVs) are one of the forms of genomic variation that are important in evolutionary processes, involved in determining population diversity and in the etiopathogenesis of certain diseases. We hypothesize that Lithuanian clean-up workers of Chernobyl nuclear disaster (LCWC) who survived and are of relatively good health, might have a unique genomic variation, that might have influenced withstanding of high ionizing radiation doses and adaptation.

Methods. LCWC (93) completed the questionnaire which includes health, lifestyle, pedigree and Chernobyl catastrophe clean-up related information. Each individual was examined by clinical geneticist. Microarray genotyping (~700 000 SNPs) was performed on 93 DNA samples extracted from peripheral venous blood of LCWC. For primary data analysis, the Illumina GenomeStudioâ□¢v2.0 Software was used. Genotyping data was called for CNVs by QuantiSNPv2.0 Software. Statistical analysis of CNV characteristics was performed and Mann-Whitney-Wilcoxon or Pearson *ï*□^{2} tests were applied (Î±=0.05). Results were compared with CNV characteristics of the control group - 286 self-reported healthy individuals from the general Lithuanian population [1].

Results. In total, there were 216 autosomal high-confidence CNVs identified in 81.8% of LCWC. 167 different CNVs were identified after the frequency analysis. The length of the CNVs was in 2.3 Kbp - 13.3 Mbp range with the mean size of 766.8 Kbp, median - 271 Kbp, and mode - 896 Kbp. The distribution of CNV size in the LCWC group and Lithuanian population differed significantly ($p=1.05*10^{-21}$). The average number of CNVs per person was 3 in LCWC group. It significantly differed from the Lithuanian population average of 9 CNVs per person ($p=2.36*10^{-5}$). Moreover, frequency of deletions (34.3%) and duplications (65.7%) differed in LCWC group (p=0.03), as well as in comparison with Lithuanian population where 50.3% deletions and 49.7% duplications were identified ($p=6.09*10^{-6}$).

Conclusions. More than 30 years after the tragedy, we provide basic CNV characteristics of the LCWC for the first time. CNV characteristicsâ□□ comparison with the Lithuanian population shows significant differences. These primary results suggest the unique genetic variation structure in LCWC group and point towards further detailed CNV analysis.

The study meets ethical standards. Informed written consents of all study participants are obtained.

This project has received funding from the Research Council of Lithuania (LMTLT), agreement No. S-MIP-20-35.

References:

[1] Urnikyte, A et al. (2016). CNV analysis in the Lithuanian population. BMC genetics, 17(1), 1-8.

PrgmNr 2135 - Differentiation in Olfactory Receptor Genes in Worldwide Populations

View session detail

Author Block: **S. Lee**[1], M. E. Phipps[2], Q. Ayub[3]; [1]Jeffrey Cheah Sch. of Med. & Hlth.Sci., Monash Univ. Malaysia, Sunway City, Selangor, Malaysia, [2]Monash Univ. Malaysia, Sunway City, Selangor, Malaysia, [3]Monash Univ. Malaysia, Bandar Sunway, Malaysia
Disclosure Block: **S. Lee:** None.
Olfaction is an ancient, evolutionarily critical sense important for detecting chemical signals in our environment and mediating our safety, nutrition, sensation of pleasure and general well-being. Odorant detection begins with the binding of a plethora of odorant molecules to olfactory receptors. To date, olfactory receptor genes constitute the largest multigene family in the human genome. Yet, much is unknown about this gene family and itsâ variants in underrepresented populations. We compiled a list of 879 olfactory receptor genes and identified their associated variants, using Human Genome Diversity Project (HGDP) data with the inclusion of data we derived from whole-genome sequencing of 16 *Orang Asli*, who represent indigenous communities in Peninsular Malaysia. We performed database searches in GTEx, Ensembl, GWAS Catalog and GeneCards to investigate these variants with their associated genes, phenotypes, traits and pathogenicity predictions. From calculations of delta derived allele frequencies (âDAF) and F_{ST}, we identified several biallelic SNPs of high impact (rs2647574, rs16930998, etc.) and over 150 SNPs of moderate impact (missense variants) that showed high differentiation in HGDP populations. The comparison of variants frequencies between the HGDP and OA revealed that rs2647574C>T in *OR51Q1* appeared to be under strong selection pressure in OA. The âwildtypeâ gene is expressed in testis, whole blood and brain. The T allelic variant is strongly selected in the OA investigated, with a frequency of 0.9375, which is substantially higher than what has been reported in populations represented in the 1000 Genomes Project described in Ensembl. Based on the single-tissue eQTLs data in GTEx, this variant decreases gene expression in testis. We continue to investigate this and other variants of interest in the olfactory receptor genes.

PrgmNr 2136 - Evaluating the accuracy of genotype imputation in the MHC region in selected African populations

View session detail

Author Block: **R. Nanjala**[1,2], N. Mulder[2], M. Mbiyavanga[2], S. Hashim[1,3], S. de Villiers[1,3]; [1]Pwani Univ., Kilifi, Kenya, [2]Univ. of Cape Town, Cape Town, South Africa, [3]Pwani Univ. BioSci.s Res. Ctr., Kilifi, Kenya

Disclosure Block: **R. Nanjala:** None.

Genome wide association studies (GWAS) traditionally use genotyping arrays to genotype large sets of individuals and thus determine which Single Nucleotide Polymorphism (SNPs) are significantly overrepresented in the cases compared to the controls and in this way determine association with disease. Genotyping arrays are cheaper than sequencing but only measure a portion of selected SNPs across the genome. To increase the number of SNPs available, one can use a reference panel of whole genome sequence data from related populations to impute SNPs from those on the array. Some regions in the human genome such as the Major Histocompatibility Complex (MHC) are highly variable and thus difficult to impute. The MHC region in humans has been associated with autoimmune and infectious diseases, adaptive and innate immune responses, and adverse responses to organ transplantation. The aim of this study is to therefore evaluate the accuracy of MHC imputation especially in African populations as they have high diversity and an extended linkage disequilibrium. The study sets will be selected from the Gambian individuals within the Gambian Genome Variation Project and simulated data from the South African population. The reference dataset will be chosen from the whole 1000 Genome population, the African sub-population within the 1000 Genome population, the Gambian sub-population within the 1000 Genome population and the H3Africa reference panel through their imputation service. Human Leukocyte Antigen (HLA) typing will be done using the OptiType tool. HLA alleles will be imputed from SNP data using HIBAG, SNP2HLA, CookHLA and Minimac4. The assessment metrics will be concordance rate, squared Pearson correlation coefficient and call rate. It is anticipated that the most appropriate software and reference panel for MHC imputation in African populations will be identified. The study will also be able to determine whether genotyping array influences accuracy and efficiency of MHC imputation.

PrgmNr 2137 - Genetic diversity in host immune genes involved in responses against intestinal parasitic infections in the Orang Asli community in Malaysia

View session detail

Author Block: **H. Kumar**[1], M. E. Phipps[2], C. S. Chuen[1], M. Ashaie[1], Y. A. Lim[3], Q. Ayub[4]; [1]Sch. of Sci., Monash Univ. Malaysia, Bandar Sunway, Malaysia, [2]Jeffrey Cheah Sch. of Med. and Hlth.Sci., Monash Univ. Malaysia, Bandar Sunway, Malaysia, [3]Faculty of Med., Univ. Malaya, Kuala Lumpur, Malaysia, [4]Monash Univ. Malaysia, Bandar Sunway, Malaysia
Disclosure Block: **H. Kumar:** None.
The â□□Orang Asliâ□□ (OA) indigenous tribes constitute 0.6% of the Malaysian population. They reside in Peninsula Malaysia and are generally classified as Negrito, Senoi and Proto-Malays. Among the Negrito group, the rural hunter-gatherer Jahai population has a high incidence of intestinal helminthic infections (68% of the samples; n=28 collected were positive for at least 1 infection.). Using whole exome sequencing, we examined genetic variation in innate and adaptive immune response genes that modulate outcome of helminthic infections in OA tribes. We also characterized variation in these genes using the Human Genome Diversity Project (HGDP) dataset, comprising of whole genome sequences of individuals belonging to 54 indigenous populations from seven geographical regions across the world. The analysis highlights variants in interleukin-7 (*IL-7*) and interleukin-34 (*IL-34*) genes that are highly differentiated between African and non-African populations and their implications for intestinal helminthic infections in the OA.

PrgmNr 2138 - Genomic and ancestral variation underscores the severity of COVID-19 disease presentation

View session detail

Author Block: **R. Das**[1], P. Upadhyai[2], G. Suresh[1], R. Parit[2]; [1]Yenepoya Res. Ctr., Yenepoya (Deemed to be Univ.), Mangalore, India, [2]Kasturba Med. Coll., Manipal Academy of Higher Ed., Manipal, India
Disclosure Block: **R. Das:** None.
Introduction: The coronavirus disease (COVID-19) is characterised by a wide spectrum of clinical manifestations ranging in severity from asymptomatic to symptomatic with mild, moderate or severe disease presentations. COVID-19 susceptibility, severity and recovery has shown discernible variability across the world. In this study we sought to identify novel SNPs that show significant frequency variation between asymptomatic versus severely affected COVID-19 patients, employing genotyping dataset generated by the AncestryDNA COVID-19 host genetic study. We further assessed the ancestral affiliations of the COVID-19 patients by combining their genetic data with ancient and modern-day human genome data.**Method**: We categorized patients into five categories denoting the acuteness of manifestation, namely asymptomatic, mild, moderate, severe and unknown, based on self-reported responses. Genome-wide association study (GWAS) was performed in PLINK v1.9 between asymptomatic individuals (controls) and severely infected patients (cases). The ancestral affiliations were determined using Principal Component Analysis (PCA), ADMIXTURE v1.3 and qpAdm algorithm implemented in AdmixTools v5.1.**Results and Discussion**: Our data revealed striking genomic differences between asymptomatic and acutely symptomatic COVID-19 patients. We identified 33,321 novel genetic variants including those associated with pathways regulating host immunity, such as innate and adaptive immune system, interferon regulated antiviral defence, interleukin signalling based immune responses, the complement cascade and known COVID-19 comorbidities, such as, hypertension and type 2 diabetes. Further, variants modulating drug responses to anti-retroviral agents, such as tenofovir, raltegravir and efavirenz were found to vary significantly between asymptomatic and severe patient groups. Using PCA, Admixture and linear regression models we showed that asymptomatic individuals contain significantly larger fractions of Bronze and Iron Age British and ancient Central European ancestries such as Bell Beaker and Corded Ware (Yamnaya related), which introduced ANE and EHG ancestries to Europe. Consistent with this, ancestry analysis using qpAdm algorithm revealed that asymptomatic individuals possess discernibly higher proportions of Ancestral North Eurasian (ANE)/Eastern Hunter Gatherer (EHG) ancestry and lower fractions of Western Hunter Gatherer (WHG) ancestry, while severely symptomatic patients have higher fractions of WHG and lower ANE/EHG ancestral components, thereby delineating the putative ancestral differences between the two groups.

PrgmNr 2139 - Gut microbiome analyses of ancient individuals, so called â□□Jomonâ□□, lived in Japanese archipelago

View session detail

Author Block: **L. Nishimura**[1,2], A. Tanino[3], A. Ajimmoto[4], K. Inada[5], T. Katsumura[3], M. Ogawa[3], K. Koganebuchi[6], D. Waku[7], R. Sugimoto[1], M. Kumagai[8], H. Oota[6], I. Inoue[1,2]; [1]Natl. Inst. of Genetics, Mishima, Japan, [2]The Graduate Univ. for Advanced Studies (SOKENDAI), Mishima, Japan, [3]Kitasato Univ., Minato, Japan, [4]Wakasa History Museum, Wakasa, Japan, [5]Hitachinaka City Archaeological Res. Ctr., Hitachinaka, Japan, [6]Univ. of Tokyo, Bunkyo, Japan, [7]Tokyo Univ. of Agriculture, Setagaya, Japan, [8]Natl. Agriculture and Food Res. Organization, Ibaraki, Japan

Disclosure Block: **L. Nishimura:** None.

Recent advancements of sequencing technologies allow facilitating ancient DNA analyses. Those ancient DNA studies have revealed not only ancient peopleâ□□s characteristics but also the ancient microbesâ□□ ones. Those microbial data may provide information about the ancient peopleâ□□s behaviors. Especially, paleofeces or coprolites are beneficial to infer diet at the ancient time. Also, we can assume the health conditions through detecting pathogenic microbiomes in the gut, which have a significant impact on our health and disease. However, there are no studies using coprolites of ancient people in Japan, and the microbial contents are unknown. Here, we utilized four paleofeces of Jomon individuals who lived in the Japanese archipelago around 16,000 to 2,900 years ago for metagenomic analyses. We conducted whole genome sequencing with the NovaSeq6000 system and aligned those obtained reads to known reference sequences. We detected several pathogenic viral sequences, such as the Human herpesvirus 2. Also, the alignment results indicated some viruses coexisted with their host species, which also exist in the modern human. It means the host-viral relationships have continued for more than thousands of years. Moreover, we detected some eukaryotic sequences, such as salmon presumably derived from food remnants. Our results indicate that the paleofeces of ancient Jomon people are helpful to understand host-viral evolutional processes and diet at the Jomon periods.

PrgmNr 2140 - Hematological, epigenetic and genetic adaptations in Tibetans

View session detail

Author Block: **N. Basak**[1,2], J. Castillo-Fernandez[3], T. Norboo[4], L. Kumar[1], M. S. Mustak[5], J. T. Bell[3], K. Thangaraj[1,2,6]; [1]CSIR-Ctr. for Cellular and Molecular Biology, Hyderabad, India, [2]Academy of Scientific and Innovative Res., Ghaziabad, India, [3]Dept. of Twin Res. & Genetic Epidemiology, King's Coll. London, London, United Kingdom, [4]Ladakh Inst. of Prevention, Leh, India, [5]Dept. of Applied Zoology, Mangalore Univ., Mangalore, India, [6]DBT-Ctr. for DNA Fingerprinting and Diagnostics, Hyderabad, India
Disclosure Block: **N. Basak:** None.
Tibetans from high altitudes have been studied in great detail for genetic and physiological adaptations. However, the adaptation pattern of high altitude Tibetans has not yet been compared to their low altitude counterparts. We studied two groups of Tibetans; one from high altitudes (â¥4500 meters above sea level (masl)), and another from low altitudes (~850 masl) in India, who share common genetic ancestry, to compare their hematological profile and epigenetic landscapes. Additionally, we explored signatures of genetic adaptation among them. Various hematological parameters from 79 high altitude Tibetans and 89 low altitude Tibetans were studied, comparing complete blood profiles evaluated by manual and automated hematology analyzer, respectively. ELISA was performed to measure serum erythropoietin. Whole genome bisulfite sequencing (WGBS) was carried out in 10 Tibetans (5 from each altitude) to compare their DNA methylomes. Thirty-six Tibetans (18 from each altitude) were genotyped and the data were compared with the already available data of the Han Chinese in Beijing (CHB), the closest low altitude population to the Tibetans . Within-population selection test (iHS) and cross-population selection tests (XP-EHH and Rsb) were performed to detect signatures of natural selection. Analyses were performed using GraphPad Prism, Bismark, and R. Significant differences in multiple hematological parameters, such as Hb, HCT, RBC, MCH (only in males) and MCV, were detected. Seventy-one significantly differentially methylated regions (DMRs) were revealed by WGBS analyses, showing >15% difference in DNA methylation levels (FDR TMPRSS6, *DUSP22*, *CIZ1*, *LRCOL1, PF4V1*). Logistic regression revealed that 10 DMRs in males and 6 DMRs in females were significantly associated with hemoglobin concentration. Results from iHS revealed 95 markers (p-value EPAS1, TMEM247, ATP6V1E2, RHOQ, PIGF, CRIPT, SOCS5*, TRPM3*. To conclude, our study identified significant hematological and epigenetic differences between the Tibetans from high and low altitudes that are likely due to the local adaptations.

PrgmNr 2141 - Historical Demography of Lithuania and Relationship to other Populations

View session detail

Author Block: **A. Urnikyte**[1], A. Molyte[1], E. Pranckeviciene[1], Z. Kucinskiene[2], V. Kucinskas[1]; [1]Dept. of Human and Med. Genetics, Inst. of BioMed. Sci., Faculty of Med., Vilnius Univ., Vilnius, Lithuania, [2]Dept. of Physiology, Biochemistry, Microbiol. and Lab. Med., Inst. of BioMed. Sci., Faculty of Med., Vilnius Univ., Vilnius, Lithuania
Disclosure Block: **A. Urnikyte:** None.

The analysis of geographically specific regions and the characterization of fine-scale patterns of The prehistory of the Lithuanian population and genetic relationship to other populations are poorly studied. Thus, the Lithuanian population, as an object of study, is interesting due to its partial isolation with genetic distinctiveness within the European context and with preserved ancient genetic composition. The main objects of this study was to infer demographic parameters, effective population size (*Ne*), and divergence time using highâ□□density single nucleotide polymorphism (SNP) genotyping data generated with the Illumina HumanOmmiExpressâ□□12v1.1 array in 295 individuals from the Lithuanian population and to compare our data with other populations from the Human Genome Cell Line Diversity Panel (HGDPâ□□CEPH). We also aimed to reconstruct past events between the main ethnolinguistic regionsâ□□AukÅitaitija and Å½emaitija of Lithuania. Historically, these regions probably developed as two independent Baltic tribes. Our results of *Ne* in the Lithuanian population through time demonstrated a substantial reduction of *Ne* over the 150,000-25,000 years before present (YBP). The estimated longâ□□term *Ne* of the Lithuanian population is quite lowâ□□it equals 5404, which likely is a consequence of the bottlenecks associated with the last glacial period of 25,000-12,000 YBP in Europe. The obtained divergence time estimates between the study populations are in agreement with recent studies. The reconstructed past events in AukÅitaitija and Å½emaitija showed significant differences between these two regions of Lithuania.This study is a part of the ANELGEMIA project, which has received funding from the Research Council of Lithuania (LMTLT), agreement No. S-MIP-20-34.

PrgmNr 2142 - Incidence of patients with methylenetetrahydrofolate reductase (MTHFR) gene mutations in a Tunisian population cohort

View session detail

Author Block: **R. Abdelhedi**[1], B. Abdelmoula[2], R. Louati[3], N. Kharrat[1], A. Rebai[4], N. B. Abdelmoula[2]; [1]Ctr. de Biotechnologie de Sfax, Sfax, Tunisia, [2]Med. Univ. of Sfax, Sfax, Tunisia, [3]FMS, Sfax, [4]centre de Biotechnologie de Sfax, Sfax, Tunisia

Disclosure Block: **R. Abdelhedi:** None.

Objective :Methylenetetrahydrofolate reductase (MTHFR) plays a central role in folate-dependent homocysteine metabolism, and severe enzyme deficiency results in elevated plasma homocysteine concentration, which is a risk factor for many pathologic conditions. Three clinically important mutations of MTHFR: C677T, A1298C, and T1317C are reported to be associated with various pathological conditions. These variants have an impact on enzyme function: 677T is affecting the catalytic and 1298C the regulatory MTHFR domain. This study was conducted to evaluate the frequency of the C677T MTHFR polymorphism in a Tunisian population cohort.Material and Methods:Our study included 85 Tunisian individuals. After genomic DNA, PCR amplification was performed using forward and reverse primers at exon 4. Genotyping for the MTHFR 677C>T polymorphism was performed by restriction digestion of PCR products with HinfI. The C677T mutation introduces a new HinfI restriction site, which results in the digestion of the 110 bp amplicon into 87 and 23 bp fragments. Results: The C677T MTHFR polymorphism was detected in 12.5 % of our population study. Eight cases were heterozygous and only one was homozygous. Conclusion: The frequency of the MTHFR 677T allele varies substantially in different regions of the world and among ethnic groups. The mean of these frequencies ranges from 27,2% to 46,2%.Conclusion:Our results of this study on MTHFR polymorphism enhance the variability of this gene worldwide and can serve as a basis for further association studies on the role of MTHFR mutation in the susceptibility of different multifactorial diseases.

PrgmNr 2143 - Population-specific adaptation to malaria infection in endemic regions of Asia

View session detail

Author Block: **E. Gusareva**[1,2,3], P. A. Lorenzini[1,3], N. A. B. Ramli[1,2,3], A. G. Ghosh[1,3], H. L. Kim[1,2,3]; [1]Asian Sch. of the Environment, Nanyang Technological Univ., Singapore, Singapore, [2]Singapore Ctr. for Environmental Life Sci. Engineering, Nanyang Technological Univ., Singapore, Singapore, [3]The GenomeAsia 100K Consortium, Singapore, Singapore
Disclosure Block: **E. Gusareva:** None.
Objective. Malaria-causing parasites (*P. falciparum, P. vivax, P. malariae, and P. ovale*) appeared in human population about 100 thousand years ago (kya) and expanded into large territories in the last 10 kya. At present, there are many human populations across Asia routinely exposed to malaria infection. Evolutionary mechanisms of adaptation to malaria are understudied in Asian endemic regions despite a high prevalence of this infection. **Methods.** By implementing three metrics (Integrated Haplotype Score - iHS, Cross-Population Extended Haplotype Homozygosity analysis - XP-EHH, and Population Branch Statistic - PBS), we have exhaustively screened for footprints of natural selection in a whole-genome sequencing dataset generated from 907 healthy individuals. Eight population groups living in malaria endemic regions across Asia (Tibeto-Burmans, Temuan-Senoi, Eastern Indonesians, Mainland Southeast Asians, Indo-Europeans, Malaysian Negritos, Melanesians, Indian Austroasiatic) and one African (Yoruba) group were compared with two non-endemic groups (Europeans and Mongols). The significance of the statistic metrics was assessed through percentile rank of the genome-wide SNPs in each population group. **Results.** We report 285 genes under positive selection in malaria-endemic populations. The identified selection signals were robust across the three statistical methods. The smallest number of malaria adaptive genes was identified in nomadic hunter-gatherers Malaysian Negritos (18 genes), while the highest in the Indian Austroasiatic group (77 genes), which is characterized by one of the highest endemicity. Interestingly, most of the identified genes (82%) were found to be under selection in a single population group, while adaptive genes shared across populations were rare. This is likely due to the independent adaptation history in different endemic populations. The most concordant adaptive genes were found to be *PTPRD*, *DPP6*, *FHIT*, *CSMD1*, *CTNNA3*, *ARHGEF7*, and *RBFOX1*. The gene ontology (GO) analysis of the 285 adaptive genes (including the most concordant ones) highlighted functional processes related to the nervous system development. The GO:0007399GO term was found to be the most over-representative (*p*-value = $1.6*10^{-7}$); 26.3% of the genes (75 out of 285) were annotated to this term. **Conclusions.** The identified adaptive genes could be related to cerebral malaria and may reduce the inflammatory response and the severity of malaria symptoms. Remarkably, our novel population genomic approach identified population-specific adaptive genes against malaria infection without the need for patient samples or individual medical records.

PrgmNr 2144 - Positive selection to altitude and phenotypic associations in Papua New Guinean highlanders

View session detail

Author Block: **M. AndrÃ©**[1], N. Brucato[2], M. Leavesley[3,4], M. Mondal[1], F-X. Ricaut[2]; [1]Estonian Biocentre, Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [2]Laboratoire Ã volution and DiversitÃ© Biologique (EDB UMR5174), UniversitÃ© de Toulouse Midi-PyrÃ©nÃ©es, CNRS, IRD, UPS, Toulouse, France, [3]Strand of Anthropology, Sociology and Archaeology, Sch. of Humanities & Social Sci., Univ. of Papua New Guinea, National Capital District, Papua New Guinea, [4]ARC Ctr. of Excellence for Australian Biodiversity and Heritage, Coll. of Arts, Society and Ed., James Cook Univ., Cairns, Australia
Disclosure Block: **M. AndrÃ©:** None.
Papua New Guinea not only has one of the oldest continuous human occupations outside Africa, but its highlands have been permanently inhabited for 20 thousand years. Previous studies have shown that Papua New Guinean highlander groups display phenotypic variation potentially linked to altitude. These results suggest that their genomes might display signatures of positive selection as observed in other high-altitude populations. Although high altitude has been one of the strongest environmental pressures for human populations for thousands of years, until now no signals of positive selection to altitude have been found in Papua New Guinean highlanders. This might be explained by the absence of genomes of Papua New Guineans living at altitudes higher than 1,500 meter above sea level. We hypothesized that the genomes of Papua New Guinean highlanders might display signatures of positive selection to altitude. We performed Cross Population Extended Haplotype Homozygosity (XPEHH) and Population Branch Statistics (PBS) selection tests using 49 newly sequenced genomes of unrelated Papua New Guinean highlanders living between 2,300 and 2,800 meter above sea level. Once genomic regions under selection were identified, we used CLUES, an approximate full-likelihood method for inferring selection, to detect the SNP most likely to drive selection in each region. We then conducted an association study between the SNPâ□□s and 13 different phenotypes found to differ between lowlanders and highlanders. Our preliminary results provide new insight on the genetic adaptation of Papua New Guinean highlanders to a hypoxic environment.

PrgmNr 2145 - TogoVar; providing integrated genomic information of Japanese population

View session detail

Author Block: **L. Toyo-oka**[1], N. Mituhashi[2], M. Kawashima[1], T. Katayama[3], S. Kawashima[4], T. Takagi[1]; [1]Japan Sci. and Technology Agency, Chiyoda-ku, Japan, [2]Res. Organization of Information and Systems, Kashiwa, Japan, [3]Database Ctr. for Life Sci., Kashiwa, Chiba, Japan, [4]Res. Organization of Information and Systems, Chiba, Japan
Disclosure Block: **L. Toyo-oka:** None.
Considering population-specific information when interpreting the result in genomic studies is important. Many genetic researches using Japanese population have been conducted due to its uniqueness for having low genetic heterogeneity. However, allele frequency information of Japanese population was scattered in each website of the research projects. The Genome Aggregation Database (gnomAD) from Broad Institute and the Allele Frequency Aggregator (ALFA) dataset from dbSNP have contributed to genetic research with its allele frequency data among several populations generated by consistent analysis protocol, however, the sample number of Japanese population in the gnomAD is only seventy-six and the sample number of Japanese in ALFA is unknown due to merging to the East Asian population. Thus, we made and have run a web service, TogoVar (https://togovar.biosciencedbc.jp) database to aggregate and provide allele frequency information among Japanese population since 2018. TogoVar included three datasets, JGA-SNP from SNP chip data, JGA-NGS from whole-exome sequence data, and GEM-J WGA from whole-genome sequencing (WGS) data whose sample numbers are 183,884, 125, and 7,609, respectively. The WGS data of GEM-J WGA dataset are deposited in our NBDC Human Database and AMED Genome group sharing Database. The collaborative work between Tohoku Medical Megabank Organization (ToMMo), BioBank Japan, and RIKEN conducted joint-calling and provided the result to TogoVar. Because the deposited data of Japanese is collected from various parts of Japan, it leads to less regional bias on the allele frequency information and its huge sample number of WGS data helps to find rare variants. Two other Japanese allele frequency database information, HGVD and ToMMo 4.7K JPN, ExAC, and ClinVar database information are imported. Variant Effect Predictor is utilized for adding annotation information and the data is converted to Resource Description Framework data. To collect more comprehensive publications of each variant, RDF converted data from PubTator Central and LitVar information via web API are merged. The APIs of TogoVar are provided in public to be easy to download information of multiple variants programmatically.In the poster, the recent achievements including an advanced search function to query complicated conditions and gene view aggregating information on each gene will be presented.

PrgmNr 2146 - Variations in genes associated with recurrent miscarriages in the HGDP CEPH dataset

View session detail

Author Block: **Q. Ayub**[1], M. Pulijala[1], K. P. Song[1], B. Silvia[2], V. Colonna[3]; [1]Monash Univ. Malaysia, Bandar Sunway, Malaysia, [2]Univ. of Campania, Caserta, Italy, [3]Inst. of Genetics and Biophysics, Natl. Res. Council, Naples, Italy
Disclosure Block: **Q. Ayub:** None.
Human reproduction is highly inefficient and appropriately 30% of all human conceptions result in live birth. Miscarriage, also known as spontaneous pregnancy loss before 24 weeks of gestation, occurs in 10-15 % of all pregnancies. The most common genetic cause of miscarriage is chromosomal aneuploidies. Genome-wide analysis and manual curation identified 972 genes associated with recurrent miscarriages or stillbirths. We conducted an *in silico* analysis to catalogue genetic variations in these 972 genes using the publicly available whole genome sequenced Human Genome Diversity Project Centre dâ□□Etude du Polymorphisme Humain (HGDP-CEPH) dataset. The analysis identified a total of 2.071,033 variants in these genes, which were further annotated by Ensembl VEP to classify impact of variants across the seven regional population groups in the HGDP-CEPH dataset. We examined the overall distribution of variants in these genes and assessed their impact. No high impact variants were found, but there were several moderate impact missense mutations that were characterized using SIFT and PolyPhen. SIFT highlighted 37% as deleterious whereas PolyPhen characterized 25% as probably damaging mutations in this apparently healthy adult population sample. Further analysis of these genes in healthy humans will aid in understanding the mutation spectrum and pathways involved in miscarriages and early human development.

PrgmNr 2147 - Classifying uncertainty and impact of uncertainty on decision-making: How do medical students handle uncertainty in prenatal exome sequencing?

View session detail

Author Block: **J. E. Klapwijk**[1,2], M. G. Polak[2], K. E. M. Diderich[1], M. I. Srebniak[1], H. T. BrÃ¼ggenwirth[1], S. Lou[3], I. Vogel[3,4,5], K. Dijkstra[2], S. R. Riedijk[1]; [1]Dept. of Clinical Genetics, Erasmus Med. Ctr., Rotterdam, Netherlands, [2]Dept. of Psychology, Ed. & Child Studies, Erasmus Univ. Rotterdam, Rotterdam, Netherlands, [3]Ctr. for Fetal Diagnostics, Aarhus Univ. Hosp., Aarhus, Denmark, [4]Dept. of Clinical Med., Aarhus Univ., Aarhus, Denmark, [5]Dept. of Clinical Genetics, Aarhus Univ. Hosp., Aarhus, Denmark

Disclosure Block: **J.E. Klapwijk:** None.

Introduction. With prenatal Exome Sequencing (ES), a higher diagnostic yield is possible. However, the number of uncertain results can also increase. It is important to elucidate what effect uncertainty has on decision-making and whether intolerance of uncertainty (IU) may impact this decision-making. This study is part of a broader and ongoing (international) research line aimed to optimize prenatal ES-centred healthcare.

Method. Based on literature and interviews with healthcare providers (HCPâ□□s) and patients, our international multi-disciplinary team identified and defined ten uncertainty types (e.g. limited prenatal phenotype presentation, possible incomplete result) in prenatal ES. Three of these uncertainty types (pathogenicity/variants of unknown significance, incidental findings, and variable expression) were used to develop eight vignettes that were then presented to 5th year medical students who just finished a prenatal genetics course (N = 51). This enabled us to pilot-test our vignettes. These students indicated per vignette whether they would report the ES result to the patient, how certain they were of their choice, and they ranked the vignettes on perceived uncertainty.

Results. Based on the uncertainty rankings, vignettes were categorized into low, moderate, and high uncertainty results. Highly uncertain results were reported less often (59%) than results low in uncertainty (98%), p **Discussion and Conclusion.** This study demonstrated that the level of perceived uncertainty influenced decision-making as well as the experience of uncertainty with these choices. Intolerance of uncertainty however did not affect studentsâ□□ decision-making. Differences in decision-making may lead to differences in care, which may not always be necessary or desirable. Moving forward, this study should be extended to experienced HCPâ□□s. It would also be relevant to investigate this paradigm further including other centres and countries. Classifying uncertainty promotes development of a common language. Gaining insight into impact of uncertainty on (HCP) decision-making will help us understand how we deal with uncertainty elicited by prenatal ES.

PrgmNr 2148 - Population biobank participantsâ response to a range of genomic results reported

View session detail

Author Block: **L. Leitsalu**[1], A. Reigo[1], M. Palover[1], T. Nikopensius[1], K. LÃ¤ll[1], K. Krebs[1], S. Reisberg[2,3], R. MÃ¤gi[1], M. Kals[1], H. Alavere[1], M. NÃµukas[1], A. Kolk[4], I. Normet[5], M-L. Tammesoo[1], E. KÃ¤Ã¤rik[6], M. Puusepp[1], K. Metsalu[1], A. Allik[7], L. Milani[1], K. Fischer[1,6], N. TÃµnisson[1,8], A. Metspalu[1,9]; [1]Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [2]Inst. of Computer Sci., Univ. of Tartu, Tartu, Estonia, [3]Software Technology and Applications Competence Ctr., Tartu, Estonia, [4]Inst. of Clinical Med., Faculty of Med., Univ. of Tartu, Tartu, Estonia, [5]Family Med. Ctr. of Medicum, Tallinn, Estonia, [6]Inst. of Mathematics and Statistics, Univ. of Tartu, Tartu, Estonia, [7]Estonian Res. Council, Tartu, Estonia, [8]Tartu Univ. Hosp., Tartu, Estonia, [9]Inst. of Molecular and Cell Biology, Univ. of Tartu, Tartu, Estonia
Disclosure Block: **L. Leitsalu:** None.

Return of results (ROR) to biobank participants is considered a good approach for participant engagement while also responding to the interest towards genetic information expressed by the general public. Estonian Biobank is a population biobank with 200 000 participants. We piloted an approach of offering individual results to participants expressing interest through participant portal. Participants were offered a report delivered to them at a 30 minute face-to-face counselling session. The report included polygenic risk scores (PRS), pharmacogenomics, moderate and high impact genetic findings. Expectations and feedback were collected through the portal before ROR, shortly after ROR and six months later.

Over two years 2957 participants received results from the biobank. Overall, participants tended to feel calm, relaxed and content, with a marginal minority reporting to feel worried, tense and upset. There was a significant increase in positive feelings reported after ROR, and conversely a decrease in uncomfortable feelings reported. The significant risk factor for feeling worried, tense and upset after ROR was feeling worried, tense and upset before ROR. Information received was considered valuable, understandable and not scary. Receiving high risk information had an impact on uncomfortable feelings reported after ROR only with certain findings, such as high PRS for ischemic heart diseases or specific risk variants for thrombophilia. Meanwhile, receiving information on thrombophilia, HBOC or HNPCC were associated with feelings of content. After six months, a majority considered registering for ROR as the right decision while one percent reported regret. Most commonly, participants mentioned enjoyable and understandable communication as a positive, while unmet expectations to receive more information in general or answers to specific health concerns were frequently mentioned as negative.

Overall, participants perceived the information received as valuable even when high risk reported was causing worry at first. More transparency of the plans regarding ROR and educating the public on limitations of ROR from a biobank are needed to handle unrealistic expectations. Face-to-face delivery of results was limiting the number of participants receiving results. Although this communication was valued by the participants, to improve the feasibility of offering results on a large scale, other means of communication would need to be considered. The feedback collected gives valuable insight on participantâs value judgements and on potential response to genetic risk information when the national personalized medicine initiative is implemented.

PrgmNr 2149 - Psychological burden of preimplantation genetic testing (PGT) on couples with multiple monogenic disorders and the role of genetic counselling in Saudi Arabia

View session detail

Author Block: **M. Alshehri**[1], A. Qari[1,2], M. Al-Sayed[1,2], A. Balobaid[1,2], W. AlQubbaj[1,3]; [1]Alfasisal Univ., Riyadh, Saudi Arabia, [2]Dept. of Med. Genetics, King Faisal Specialist Hosp. and Res. Ctr. (KFSH&RC), Riyadh, Saudi Arabia, [3]Dept. of Pathology and Lab. Med., King Faisal Specialist Hosp. and Res. Ctr. (KFSH&RC), Riyadh, Saudi Arabia
Disclosure Block: **M. Alshehri:** None.
Psychological burden of preimplantation genetic testing (PGT) on couples with multiple monogenic disorders and the role of genetic counselling in Saudi Arabia. **BACKGROUND** Preimplantation genetic Testing (PGT) is a technique that provides a practical alternative to prenatal diagnosis and termination of pregnancy for couples who are at substantial risk of transmitting a serious genetic disorder to their offspring. The interplay between multiple familial genetic diseases, preimplantation genetic testing (PGT) and uncertain outcomes can have a detrimental, long-lasting psychosocial impact as it can drain individuals mentally, emotionally and physically. **OBJECTIVES** As the presence of multiple genetic disorders increases the risk of yielding affected embryos, this study focused on the emotional, social, economical burdens and attitudes towards PGT and the role of genetic counselling in IVF/ PGT clinic. **METHODS** This study was performed by interviewing 31 couples by phone. The couples were presented to the clinic with two monogenic disorders and had underwent PGT between January 2009 and March 2020 at King Faisal Specialist Hospital and Research Centre (KFSHRC), Riyadh. This study focused on participantâ□□s sociodemographic background, reproductive history, types of genetic conditions, number of children (healthy/affected), number of PGT cycles undergone, and attendance at genetic counselling sessions, and their knowledge about recurrence risk and chances of having unaffected children by multiple diseases and prenatal diagnosis (PND). The data was analysed by SPSS. **RESULTS** As the PGT was provided free of charge at KFSHRC , the main concern in this study was the psychosocial impact. Several psychosocial variables were found to have a profound impact on those couples. Therefore, it is of importance to implement a tailored genetic counselling sessions for couples with multiple genetic diseases attending the IVF/PGT clinics. **Conclusion**. More measures and adjustments to the genetic counselling sessions must be implemented for patients with multiple genetic diseases in IVF/PGT clinics.

PrgmNr 2150 - Role of whole genome sequencing in detecting compound heterozygotes for single nucleotide variant and structural variant: Two illustrative patients

View session detail

Author Block: **M. Yamada**[1], T. Shima[2], H. Adachi[3], A. Noguchi[3], H. Suzuki[1], T. Takahashi[3], K. Kosaki[1]; [1]Ctr. for Med. Genetics, Keio Univ. Sch. of Med., Tokyo, Shinjuku-ku, Japan, [2]Dept. of Pediatrics, Juntendo Univ. Urayasu Hosp., Chiba, Japan, [3]Dept. of Pediatrics, Akita Univ. Graduate Sch. of Med., Akita, Japan
Disclosure Block: **M. Yamada:** None.
In the molecular diagnosis of autosomal recessive diseases, identification of both pathogenic alleles and their parental origin is essential for accurate recurrence risk counseling. Exome analysis, the current mainstay of genetic testing, can reproducibly detect single nucleotide variants whereas the method can offer limited ability in detecting structural variants. When a patient with an autosomal recessive disorder is a compound heterozygote for single nucleotide variant and structural variant, exome analysis often fails to detect the latter variant. Here, we report two such compound heterozygotes diagnosed successfully by whole-genome sequencing and recurrence risk of 25% were given to both parents. 1) 2-year-old male who presented with severe neonatal asphyxia, severe intellectual disability, and intractable seizures. He was a compound heterozygote for *SMPD4* pathogenic alleles: hemizygous (apparently â□□homozygousâ□□) nonsense variant NM_017951.4 c.832C>T, p.(Arg278*) and 56-kb deletion spanning *SMPD4*, *MZT2B*, and *TUBA3E*. 2) 19-days-old male who presented with congenital pulmonary hypoplasia, fetal hydrops, and cerebellar hypoplasia. He was a compound heterozygote for *SLC25A46* pathogenic alleles: hemizygous (apparently â□□homozygousâ□□) splice acceptor site NM_138773.2 c.385-1G>A and 80-kb deletion spanning *SLC25A46* and *TMEM232*. If deletions had not been detected, *de novo* occurrence of the single nucleotide variant or uniparental isodisomy might have been considered. In both situations, recurrence risk would have been zero. These two examples illustrate importance of pursuing structural variation when evaluating potentially pathogenic single nucleotide variants.

PrgmNr 2151 - Stigma-power around Fragile X Syndrome: a tale of a royal family and a community in a rural village in Cameroon

View session detail

Author Block: **K. Kengne Kamga**[1], N. Syntia Munung[1], S. Nguefack[2], A. Wonkam[3], J. Devries[1]; [1]Univ. of Cape Town, Cape Town, South Africa, [2]Univ. of Yaounde 1, Yaounde, Cameroon, [3]Univ Cape Town, Cape Town, South Africa

Disclosure Block: **K. Kengne Kamga:** None.

Abstract

Fragile X Syndrome (*FXS*) is a neurogenetic condition that significantly impacts the lives of affected individuals and their families due to its association with intellectual disability (ID) and stigma. In this paper, we present the findings of an ethnographic study in the community of a patient who received the genetic diagnosis for *FXS* in Cameroon. This study builds on data from 28 participants of a royal family and 58 from the community who participated in 20 in-depth interviews and nine focus group discussions.

We identified two types of stigma in this community: public stigma directed towards the royal family and associative stigma experienced by royal family members. We outline the stereotyping labels used for the family and its children with Fragile X Syndrome and describe the stigma-power dynamic between the community members and the royal family. First, most villagers use less stigmatizing terms to addressing *FXS* children from the chieftaincy because of their position in society. Secondly, due to their social position, the royal family uses their status to negotiate marriages with community members. From these observations, we can suggest that the primary role of stigma in this community is to keep people away from *FXS* and keep them down through domination and exploitation.

We advocate that other researchers examine if the same pattern exists in other inheritable forms of ID and conduct more qualitative research on FXS in Africa.

keywords: Fragile x syndrome, stigma-power, Qualitative study, Cameroon

PrgmNr 2152 - The impact of clinically relevant CNVs in the general population - the health consequences and personalized management of undiagnosed adult CNV carriers in the Estonian biobank

View session detail

Author Block: **M. NÃµukas**[1], M. Palover[1], M. Lepamets[1], L. Pizzo[2], K. Lehto[1], A. Reigo[1], H. Alavere[1], L. Leitsalu[1], I. Xenarios[3,4], M. Metspalu[5], A. Reymond[6], S. Girirajan[2], N. TÃµnisson[1,7], K. MÃ¤nnik[3]; [1]Estonian Biobank, Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [2]Biochemistry & Molecular Biology, and Anthropology, Pennsylvania State Univ., University Park, PA, [3]Hlth.2030 Genome Ctr., Geneva, Switzerland, [4]Lausanne Univ. Hosp., Lausanne, Switzerland, [5]Univ. of Tartu, Tartu, Estonia, [6]Univ Lausanne, Lausanne, Switzerland, [7]Dept. of Clinical Genetics in Tallinn, Tartu Univ. Hosp., Tartu, Estonia

Disclosure Block: **M. NÃµukas:** None.

The role of CNVs is well established in children with neurodevelopmental disorders (NDDs). However, more research is needed towards systematic understanding on how: i) CNVs affect health in the adult population and ii) to responsibly disclose findings to CNV carriers at high genetic risk for complex disorders.

We screened the cohort of Estonian Biobank (EstBB; n=132,770) for 81 recurrent CNV regions associated with susceptibility to NDDs. In the first stage, we used the â□□genotype-firstâ□□ approach to fetch health data from the EstBB, linked electronic health registries (EHRs) and mapped each ND-CNV with their co-occurring disease traits. In the second stage, we selected 10 ND-CNVs as a paradigm to return of genetic risk data and analysis of at-risk individualsâ□□ experience and the impact of disclosed genetic finding on their health support.

Our results show that the prevalence of CNVs associated with susceptibility to NDDs in the EstBB is 2.6% (n=3,404). Further prioritization of CNVs listed in the DECIPHER database suggested a population prevalence of 0.5% (n=710) for clinically well-established CNV syndromes. According to EHRs, nearly half of them have previously documented neurological or mental and behavioural problems. Notably, only 4 out of 710 are aware of their genetic diagnosis.

Our results along with reports by others confirm that CNVs associated with NDDs are cumulatively common, but still understudied health problem in general population. This work was supported by the Jacobs Foundation Research Fellowship (Dr MÃ¤nnik), Swiss National Science Foundation grant (Dr A.R) and the Estonian Research Council grant (Dr TÃµnisson).

PrgmNr 2153 - Use of MOOCs for inclusive, flexible professional development and wider public engagement in human genomics and genetics

View session detail

Author Block: **D. Nikolic**[1], R. Berkson[1], T. Creavin[1], I. Rafi[2], J. Hayward[3], D. Hipps[4], W. Evans[5], A. Middlton[1], J. Roberts[1]; [1]Wellcome Connecting Sci. (WCS), Hinxton Cambridgeshire, United Kingdom, [2]St Georgeâ□□s, Univ. of London, London, United Kingdom, [3]Shipley Med. Practice, Shipley, United Kingdom, [4]The Exchange Surgery, London, United Kingdom, [5]NHS Leeds, Leeds, United Kingdom
Disclosure Block: **D. Nikolic:** None.
Background: Human genetic and genomics knowledge and technology are becoming increasingly prevalent, creating an urgent need to integrate them into a wide range of professions, especially healthcare. Patients increasingly experience genomics as part of their daily lives; therefore, healthcare providers should be informed about the basic science and implications, so that they can provide appropriate guidance. The availability of commercial (DTC) genomic testing has increased this need. Scalable, inclusive education/training on genomics, flexible to fit professionalsâ□□ busy lives, is difficult to achieve.
Objectives: WCS produced two advanced training courses aimed at practitioners in genetic counselling (GC) and primary care (PC) to help address this need. The main aims were to provide inclusive, free and accessible courses for a global audience of healthcare professionals but also for a wider public interested in the topics. The objectives of the PC course included studying the role of genomics within primary care and management of patients with direct to consumer test results. The objective of the GC course was to raise awareness of the role of genetic counselling in the genomic era, from both professional and public points of view, presenting diverse approaches and challenges to genetic counselling in different countries around the world. Both courses addressed some of the ethical issues connected to genetic testing.
Methods: The courses were run on a massive, open, online platform (MOOC), attended by global audiences of thousands of learners. Conversation-based pedagogy was applied to harness social learning and encourage exchange among different types of experts from around the world. Asynchronous, written and lightly facilitated by PC and GC experts, they offered opportunities for learners to study at their own pace. Both courses made extensive use of realistic, clinically relevant case studies, which the learners investigated and discussed.
Results: Between 2019-21, 2 courses had 5 runs altogether. They reached nearly 12000 learners from more than 122 different countries. Emerging themes from the qualitative analysis of learnersâ□□ comments include generally high learner satisfaction, especially with provision of opportunities to explicitly integrate the theory learned on the course within their work practice, and acquisition of tools and resources for further use. Learnersâ□□ intention to apply knowledge acquired on the course in professional practice or personal life was detected, which points to the use of MOOCs as a viable option in both updating professionals and for wider public engagement in genomics and genetics topics.

PrgmNr 2154 - Novel CRISPR system with modified cas12 and nanoliposome based delivery can destroy SARS covid2 virus in human lung tissue

View session detail

Author Block: **A. Haghighatfard**; Arvin gene company,Tehran, Iran, Tehran, Iran, Islamic Republic of
Disclosure Block: **A. Haghighatfard:** None.
SARS-CoV-2 virus is a RNA virus that may cause severe respiratory disorder with unclear rate of death. Genome editing systems may target viruses with both vaccines and pharmaceutical approach. Previously we designed a CRISPR system to reduce the expression level of a LncRNA that was involved in depression and addiction in human. The system was successful to inhibit the LncRNA in both cell line and transgenic mouse.We designed and synthesized CRISPR RNAs (crRNAs) which targeting conserved viral and functional regions of SARS-CoV-2 sequence. The same deep learning analysis of previous study on LncRNAs were used to design and choose the most effective seven crRNAs that target 98 percent of all coronaviruses and have none off target sequence in human or microflora genome. Also most suitable sequence of synthetic Cas protein for CRISPR-Cas and prophylactic antiviral CRISPR (PACMAN) were choose by deep learning and protein design analysis. Synthesized CRISPR-Cas system delivered to covid19 affected human lung epithelial cells, neuronal cells and heart cells as well as SARS and covid19 positive rats for 48 hours. Then experiment group of cell lines and rats examined for degradation of viral RNA of SARS-CoV-2. Also harmlessness for host genome expression profile was assessed in control group that received only CRISPR system as well as experiment group. Expression study for viral RNA on covid19 affected human cell lines and lung, brain and heart tissue of SARS and covid19 positive rat models were conducted by using Real time PCR, flowcytometry and western blotting. Also RNA sequencing performed for analysis of gene expression changes in human genome and rat genome for cell lines and rat models respectively.Results were showed 84 to 97 percent down regulation of covid19 and SARS RNAs in different dosage of CRISPR system. Also no significant changes were detected in gene expression patterns of human cell lines and rat tissues compared with control group. Results could suggest the effectiveness of CRISPR system to attack corona viruses or maybe other viruses with no side effects on human body. In addition further analysis on cell lines that get the CRISPR system before infection by covid19 may reveal the vaccination effect of our PACMAN system.

PrgmNr 2155 - Targeted Therapies for Hereditary Peripheral Neuropathies: Systematic Review and Steps Towards a 'treatabolome'

View session detail

Author Block: **A. Lochmuller**[1], M. J. Jennings[1], A. Atalaia[2], R. Horvath[1]; [1]Univ. of Cambridge, Cambridge, United Kingdom, [2]Sorbonne UniversitÃ©, Paris, France
Disclosure Block: **A. Lochmuller:** None.
Background: Hereditary peripheral neuropathies are inherited disorders affecting the peripheral nervous system, including Charcot-Marie-Tooth disease, familial amyloid polyneuropathy and hereditary sensory and motor neuropathies. While the molecular basis of hereditary peripheral neuropathies has been extensively researched, interventional trials of pharmacological therapies are lacking. Objective: We collated evidence for the effectiveness of pharmacological and gene-based treatments for hereditary peripheral neuropathies. Methods: We searched several databases for randomised controlled trials (RCT), observational studies and case reports of therapies in hereditary peripheral neuropathies. Two investigators extracted and analysed the data independently, assessing study quality using the Oxford Centre for Evidence Based Medicine 2011 Levels of Evidence in conjunction with the Jadad scale. Results: Of the 2046 studies initially identified, 119 trials met our inclusion criteria, of which only 36 were carried over into our final analysis. Ascorbic acid was shown to have no therapeutic benefit in CMT1A, while a combination of baclofen, naltrexone and sorbitol (PXT3003) demonstrated some efficacy, but phase III data are incomplete. In *TTR*-related amyloid polyneuropathy tafamidis, patisiran, inotersen and revusiran showed significant benefit in high quality RCTs. Smaller studies showed the efficacy of L-serine for *SPTLC1*-related hereditary sensory neuropathy, riboflavin for Brown-Vialetto-Van Laere syndrome (*SLC52A2/3*) and phytanic acid-poor diet in Refsum disease (*PHYH*). Conclusion: The 'treatable' variants highlighted in this project will be flagged in the treatabolome database to alert clinicians at the time of the diagnosis and enable timely treatment of patients with hereditary peripheral neuropathies.

PrgmNr 2156 - A global omics data sharing and analytics marketplace: Case study of a rapid data COVID-19 pandemic response platform

View session detail

Author Block: **G. Lalli**[1,2], I. Farah[3,2,4], K. Van Steen[1], D. J. Baker[2,5], A. Schumacher[2]; [1]BIO3 â Lab. for Systems Med., Katholieke Univ.it Leuven, Leuven, Belgium, [2]Shivom Ventures Limited, 71-75 Shelton Street, Covent Garden, London, England, WC2H 9JQ, London, United Kingdom, [3]Div. of Psychiatry, Univ. Coll. London, London, United Kingdom, [4]Novo Nordisk Res. Ctr. Oxford, Innovation Building, Old Road Campus, Roosevelt Drive, Oxford, United Kingdom, [5]GoldenHelix Fndn., London, England, WC2N 5AP, London, United Kingdom
Disclosure Block: **G. Lalli:** None.
The sudden, rapid spread of the coronavirus (more specifically, SARS-CoV-2) that hit the world in December 2019, triggering a pandemic that is still ongoing today, has made the lack of proper coordination between public and private healthcare sectors even more evident. These deficits caused the global health system to respond too late to the initial SARS-CoV-2 outbreak, thus slowing the identification of the pathogen triggering the disease condition, as well as understanding how the host genome influences disease severity.â¯â¯

Even though an ever-increasing amount of data is continuously being generated, their use is too often limited because they present complicated access hurdles, limited data-sharing capabilities, and missing interoperability and analytics capabilities that limit their usefulness.â¯Newly generated data, however, not only incurs the problem of siloing but also shows that it does not meet adequate quality standards due to the small sample size, which limits its re-usability, a quality necessary to make the information usable and valuable.â¯

Here we describe an approach that we implemented to provide healthcare professionals with a platform to retrieve and analyse genetic data, and securely and share sensitive patient data while ensuring the deidentification of the patient. Individuals can not be identified because only algorithms can access the raw data and researchers querying the data hub do not have access to the source of the genetic data. â¯

The Shivom platform (https://www.shivom. io/), is built on a principle of open science, transparency and collaboration, and designed to give researchers a unique ecosystem to easily, securely and quickly share and analyse omics data, thereby simplifying the data acquisition and analysis processes. The platform can help the global research community to rapidly discover new preventive measures and optimise logistical solutions. Researchers can upload data and analyse it using predefined bioinformatics and AI algorithms in a trusted environment where they can store data in a secure, anonymised, and highly ecrypted manner, the latter to prevent malicious attacks or unauthorized access.â¯

Through proof-of-concept case studies, we show the necessities and potentials of our precision medicine platform, including providing the global research community with an online marketplace for rapid data-sharing, managing & analytics capabilities, and delivering AI insights that can better understand complex diseases, ageing & longevity, apart from global pandemics, at the molecular and epidemiological level.â¯â¯

PrgmNr 2157 - An atlas of associations between polygenic risk scores from across the human phenome and circulating metabolic biomarkers

View session detail

Author Block: **S. Fang**[1], M. V. Holmes[2], T. R. Gaunt[1], G. Davey Smith[1], T. G. Richardson[1,3]; [1]MRC Integrative Epidemiology Unit (IEU), Population Hlth.Sci., Bristol Med. Sch., Univ. of Bristol, Bristol, United Kingdom, [2]Clinical Trial Service Unit & Epidemiological Studies Unit, Nuffield Dept. of Population Hlth., Univ. of Oxford, Oxford, United Kingdom, [3]Novo Nordisk Res. Ctr., Oxford, United Kingdom

Disclosure Block: **S. Fang:** None.

Polygenic risk scores (PRS) are becoming increasingly popular for predicting complex traits and diseases, although they also hold the potential to develop insight into the molecular profiles of patients with an elevated genetic predisposition to diseases. We constructed an atlas of associations between 130 PRS for different traits and diseases with 249 circulating metabolic biomarkers measured using the Nightingale Health platform in up to 83,004 participants from the UK Biobank (UKB).We demonstrate the value of this atlas in exemplar by a hypothesis-free evaluation of genetically predicted associations with glycoprotein acetyls (GlycA), a marker of inflammation. In total, there were 49 PRS (using SNPs with r^2-09) and cigarettes per day (Beta=0.115, 95% CI=0.066 to 0.163, P=4.09x10^{-06}) on GlycA levels rather than the reverse direction of effect, suggesting that a causal effect of GlycA is unlikely to underlie the PRS associations identified in our analysis.Next, associations between PRS and metabolic biomarkers were generated within age tertiles to investigate the influence of contingent mediators which may undermine inference. To highlight the utility of this sensitivity analysis, we compare lipoprotein lipid profiles associated with the PRS for coronary artery disease (CAD) within subgroups with differing numbers of individuals undergoing statin therapy. For example, the CAD PRS was strongly associated with circulating apolipoprotein B (apo B) amongst participants in the lowest age tertile (statin users 5%, Beta=0.059, 95% CI=0.048 to 0.069, P-300), whereas weak evidence of association was found amongst participants in the oldest age tertile with the beta coefficient directionally opposite (statin users 29%, Beta=-0.009, 95% CI=-0.021 to 0.003, P=0.141). Given the incontrovertible evidence that circulating apo B causes CAD, these findings illustrate the detrimental impact of a complex mediator (e.g. statin medications) in distorting findings when analysing metabolic biomarkers in UKB.We envisage that our atlas of associations will help to develop a greater understanding of genetic liability to disease and metabolic profiles. All results can be interactively visualised at http://mrcieu.mrsoftware.org/metabolites_PRS_atlas.

PrgmNr 2158 - Human genetics is great: A genomic structural variant map of 945 Han Chinese individuals using long-read sequencing data

View session detail

Author Block: **J. Gong**[1], Y. Huang[1], Y. Ling[2], R. Cao[2], M. Pan[1], J. Tan[1], Y. Ma[1], J. Li[1], S. Wang[2], J. Wang[1], G. Zhang[2], L. Jin[1], F. Sedlazeck[3], S. Fan[1]; [1]Fudan Univ., Shanghai, China, [2]Shanghai Inst. of Nutrition and Hlth., Chinese Academy of Sci., Shanghai, China, [3]Baylor Coll. of Med., Houston, TX
Disclosure Block: **J. Gong:** None.
Structural variants (SVs) are a major source of genetic diversity that have strongly shaped human evolution and, among many other physiological impacts, continue to influence the differential susceptibility and responsivity of individuals and populations to disease. Here, we used Oxford Nanopore technology to sequence the genomes of 945 Han Chinese individuals and then characterized the SV diversity of this population and conducted extensive comparative analyses using our SV catalogue. Specifically, we identified 96,494 SVs, among which 51.98% have not been reported before. Genotyping analysis with a data for a separate group of 208 Han Chinese genomes confirmed that many of the discovered SVs are true genomic polymorphisms. After identifying 50 Mb and 255 Mb of SV mutation 'hotspot' and 'desert' genome regions, we detected SV mutation in the genome. A gene-set enrichment for pathways related to both carbohydrate and protein metabolism. We noted that the SV deserts have clearly experienced significantly stronger purifying selection compared to the hotspots, and the deserts harbor genes that are predicted to be 'essential' in humans. We uncovered 302 'natural knockout genes' in which more than 72% (217 out of 302) have not been reported before. Further, 228 SVs overlap and 1,099 SVs border SNPs previously associated with phenotypic and disease traits from GWAS. Finally, our dataset included 38 Mbp of sequence that is missing from the current human reference genome (GRCh38), which we supported with comparative analyses of non-human primate genomes. Our genomics and SV datasets are publicly available at https://www.biosino.org/chsvdb/, and represent a valuable resource for researchers in the evolutionary and population genetics, functional genomics, and medical communities.

PrgmNr 2159 - Measuring sensitivity of semi-automated Clingen-ACMG classification framework for CNVs through comparison with manual classification of 1437 clinical cases

View session detail

Author Block: **S. Tzur**[1], I. Maya[2,3], S. Kahana-Pasternak[2], I. Agmon-Fishman[2], C. Klein[2], I. Vustianiuk[1], V. Lamtieva[1], K. Sund[1], T. Talmy[1], R. Attali[1], R. Matar[2], L. Basel-Salmon[2,3,4]; [1]Genomic Res. Dept., Emedgene Technologies, Tel Aviv, Israel, [2]Recanati Genetics Inst., Beilinson Hosp., Rabin Med. Ctr., Petach Tikva, Israel, [3]Sackler Faculty of Med., Tel Aviv Univ., Tel Aviv, Israel, [4]Felsenstein Med. Res. Ctr., Rabin Med. Ctr., Petach Tikva, Israel
Disclosure Block: **S. Tzur:** Major Stockholder/Ownership Interest; Emedgene. Salary/Employment; Emedgene.
To assist laboratories in classifying copy number variants (CNVs), ACMG and ClinGen have introduced a quantitative, evidence-based, scoring framework divided into five sections. Emedgene Technologies has developed a semi-automatic classification tool following the frameworkâ□□s criteria (excluding literature survey and family history) to allow for rapid and effective CNV classification.In this study, we estimated the sensitivity of the framework and the automated tool in detecting pathogenic variants. We compared the toolâ□□s classifications with manual classifications of 5773 CNVs identified among 1437 unrelated samples. These samples were tested during 2020 by the clinical laboratory at the Rabin Medical Center using an Applied Biosystems CYTOSCAN 750K array.In 112/1437 cases (7.8%), an abnormal result was reported, including 37 samples (2.6%) with aneuploidy and 75 samples (5.2%) with 82 pathogenic (P) or likely pathogenic (LP) CNVs. Of the 82 P/LP variants reported by the laboratory, 60 were also classified as P/LP by the automated tool (sensitivity of 73%, CI[62,82]). In most of the remaining 22 variants, classified manually as pathogenic (and VUS by the tool), the variant did not overlap with a dosage-sensitive gene/region or failed to contain a sufficient number of genes. Of them, 19 overlapped with at least one known ClinVar pathogenic variant. In total, 96% of the variants were classified as P/LP by the framework or annotated with Clinvar known variant.The most impactful framework metric was 2A, which indicates overlap with a dosage-sensitive gene or region; it was indicated 49 times (score 1). Tags 3C (17 times, score 0.9) and 3B (11 times, score 0.45), which reflect the number of genes within the affected region, were also impactful. Other tags were applied less frequently.Overall, we found that the Clingen/ACMG framework, in combination with the ClinVar database, was sensitive for detecting the majority of pathogenic events, specifically in the case of a very large CNV or full overlap with a region of dosage sensitivity. The remaining portion of the pathogenic variants still demand a deeper manual investigation. Machine learning approaches surveying other features in an integrative manner could potentially help reduce the time of analysis of CNVs.

PrgmNr 2160 - SpliceAI algorithm identified potentially pathogenic variants in patients with hereditary hearing impairment

View session detail

Author Block: **C-Y. Tsai**[1,2], M-Y. Lo[3], C-J. Hsu[4,2], J. S. Hsu[5], C-C. Wu[2], P-L. Chen[6,1]; [1]Graduate Inst. of Med. Genomics and Proteomics, Natl. Taiwan Univ. Coll. of Med., Taipei, Taiwan, [2]Dept. of Otolaryngology, Natl. Taiwan Univ. Hosp., Taipei, Taiwan, [3]Inst. of Molecular Med., Natl. Taiwan Univ., Taipei, Taiwan, [4]Dept. of Otolaryngology, Taichung Tzu Chi Hosp., Taichung, Taiwan, [5]Natl. Taiwan Univ. Coll. of Med., Taipei, Taiwan, [6]Dept. of Med. Genetics, Natl. Taiwan Univ. Hosp., Taipei, Taiwan
Disclosure Block: **C. Tsai:** None.
Background: Hereditary hearing impairment (HHI) is a common inherited disorder. More than 200 genes have been identified as causally related to HHI, and the types of genetic variants that may lead to HHI are highly diverse. Although next-generation sequencing (NGS) technology has revolutionarily improved the genetic diagnosis of HHI, many families remain unsolved. Emerging bioinformatic tools, such as the â□□SpliceAIâ□□ algorithm (Illumina), provide an opportunity to analyze previously unexplored DNA regions and identify obscure pathogenic variants. **Method:** Forty unsolved HHI families, including 30 simplex families without confirmatory diagnoses on panel-based examination and 10 multiplex families without confirmatory diagnoses on the whole genome sequencing, were recruited. Their NGS data were subjected to SpliceAI analyses to search for causative variants. The probability (scoring from 0 to 1; the effective threshold is 0.5) of each variant to gain/lose the splicing donor/acceptor was assessed, followed by the loci predicted to cause alternative splicing. **Result:** The SpliceAI algorithm identified 5 potentially pathogenic variants in 5 families, including 3 recessive variants in 3 of the 30 simplex families and 2 dominant variants in 2 of the 10 multiplex families. For the former 3 families, two intronic variants c.6591-3T>G of *PTPRQ*, c.3894+5G>C of *OTOF*, and one synonymous variant c.3864G>A (p.A1288=) of *OTOF*, were predicted to cause alternative splicing. Identification of these 3 variants ascertained biallelic mutations in the 3 affected probands and confirmed the genetic diagnosis, with the genotypes being *PTPRQ* c.[6591-3T>G];[6591-3T>G], *OTOF* c.[3894+5G>C];[5098G>C] and c.[3864G>A];[5098G>C], respectively. For the latter two multiplex families, two intronic variants c.1183+5G>A of *GSDME* and c.1013+5G>A of *MITF*, which were predicted to cause alternative splicing, also co-segregated with the phenotype of hearing impairment in the affected family members. **Conclusion:** The SpliceAI algorithm can identify variants that may result in aberrant splicing, and help achieve genetic diagnoses in HHI families unsolved by current analytical protocols. However, subsequent functional assays are warranted to elucidate the exact mechanisms of alternative splicing for these variants.

PrgmNr 2161 - Towards a more usable database: MGeND 2021 update

View session detail

Author Block: **M. Kamada**[1], R. Ashine[1], M. Nakatsui[2], Y. Okuno[1,3]; [1]Graduate Sch. of Med., Kyoto Univ., Kyoto, Japan, [2]Yamaguchi Univ. Graduate Sch. of Med., Yamaguchi, Japan, [3]RIKEN Ctr. for Computational Sci., Kobe, Japan
Disclosure Block: **M. Kamada:** None.
The success of global genomic medicine requires data collection that takes into account genomic diversity. Therefore, we have developed and are operating the Medical Genomics Japan Database (MGeND), a database that aggregates genomic data and corresponding clinical information for the East Asian population. Our database was developed focusing on cancer, rare diseases, infectious diseases, dementia, and hearing loss. In addition, MGeND aggregates and integrates variant data across monogenic and polygenic diseases. More than 150,000 variants have been published in MGeND, which are accessed not only from Asia but also worldwide. The fact that 80% of the variants registered are only in MGeND demonstrates the importance and usefulness of data aggregation in Asian populations. In addition, user comments have been used to improve usability since its release in 2018. In 2020 and beyond, we developed the specific view for GWAS datasets, linked variants between genome assemblies (hg19 and hg38), expanded downloadable data, and developed APIs. In this presentation, we will report updates of the database a detailed statistical analysis of the registered data.

PrgmNr 2162 - Variations in Nomenclature of Clinical Variants between Annotation Tools

View session detail

Author Block: **K-J. Park**[1], J-H. Park[2], H-T. Kim[1], C-H. Park[1], Y-J. Kim[1]; [1]Samsung Changwon Hosp., Changwon, Korea, Republic of, [2]Samsung Med. Ctr., Seoul, Irwon, Korea, Republic of
Disclosure Block: **K. Park:** None.
Background: Accurate nomenclature of variants is an essential element for genetic diagnosis and patient care. The aim of this study was to investigate the difference in annotation of the clinical variants from the HGMD between ANNOVAR and snpEff, which are the most commonly used free tools in the clinical setting.6,7 Furthermore, the accuracy of the annotation tools was evaluated based on the standardized HGVS nomenclature. **Methods:** Clinical variants (n = 229,161) registered in the HGMD (professional version 2019.01) were described at the cDNA level and protein level according to the HGVS nomenclature. Multiple nomenclatures based on RefSeq transcripts were provided using both ANNOVAR and snpEff. Multiple nomenclatures from several alternative transcripts for each variant were annotated. **Results:** The concordance rates between ANNOVAR and snpEff of coding and protein variants were 93% and 73%, respectively. Discordant coding variant nomenclature was more identified in deletion-insertion (100%), duplication (52%), or insertion (17%) variants than substitution variants (1%). Based on the Human Genome Variation Society (HGVS) nomenclature, snpEff was more accurate than ANNOVAR (coding variants: 97% vs. 94%, protein variants: 98% vs. 73%). When annotating each variant with both ANNOVAR and snpEff, the accuracy was improved up to 99%. **Conclusions:** There were substantial differences of between ANNOVAR and snpEff annotations. This study suggests that simultaneous use of multiple annotation tools could decrease nomenclature errors and contribute to providing of standardized clinical reporting.

PrgmNr 2163 - Two new patients with focal dermal hypoplasia: A novel PORCN variant and insights on the differential diagnostic considerations

View session detail

Author Block: **R. Elhossini**[1], m. Abdel-Hamid[1], H. Dawoud[2], M. Aglan[1], M. ElRoby[1], E. Ashaat[1]; [1]Natl. Res. Ctr. of Egypt, Cairo, Egypt, [2]Tanta Univ., Cairo, Egypt
Disclosure Block: **R. Elhossini:** None.

Two new patients with focal dermal hypoplasia: A novel PORCN variant and insights on the differential diagnostic considerationsâ□□

Mutations in the *PORCN* gene cause an X-linked dominant condition, focal dermal hypoplasia (FDH), characterized by atrophic skin, pigmented skin lesions, several ocular and skeletal malformations. FDH is rare with around 275 cases reported so far from diverse ethnic groups. Herein, we provide the first report of two new patients with FDH from Egypt. In addition to the typical clinical manifestations of the disease, infrequently reported clinical findings in the form of broad metaphysis, bilateral short broad femurs and dermal sinus over the sacrum were seen in Patient 1 and partial fusion of labia majora was present in Patient 2. Two heterozygous protein truncating *PORCN* mutations were identified in our patients, a known nonsense (c.370C>T, p.Arg124Ter) and a novel frameshift (c.375delG, p.Ala126HisfsTer3). Segregation analyses confirmed that the two mutations were â□□de novoâ□□ and not inherited from any of the parents. Our study expands the clinical and mutational spectrum of focal dermal hypoplasia and emphasizes the importance of investigating the different body systems and organs for early management of patients.

PrgmNr 2164 - A human importin-Ã-related disorder: Syndromic thoracic aortic aneurysm caused by bi-allelic loss-of-function variants in *IPO8*

View session detail

Author Block: **A. Verstraeten**[1], I. Van Gucht[1], J. A. Meester[1], J. Rodrigues Bento[1], M. Bastiaansen[1], J. Bastianen[1], I. Luyckx[1,2], L. Van Den Heuvel[1], P-J. Guns[3], E. Fransen[1], J. Velchev[1], M. H. A. Perik[1], I. Pintelon[3], M. P. Ferla[4], A. R. Dallosso[5], Genomics England Research Consortium, J. A. Rosenfeld[6], K. Vijayakumar[7], Z. N. Al-Hassnan[8], J. Vogt[9], N. Revencu[10], I. Maystadt[11], E. M. Miller[12], K. Weaver[12], A. H. Begtrup[13], R. Maroofian[14], A. T. Pagnamenta[4], L. Van Laer[1], B. L. Loeys[1]; [1]Univ. of Antwerp/Antwerp Univ. Hosp., Edegem, Belgium, [2]Radboud Univ. Nijmegen Med. Ctr., Nijmegen, Netherlands, [3]Univ. of Antwerp, Edegem, Belgium, [4]Univ. of Oxford, Oxford, United Kingdom, [5]Southmead Hosp., Bristol, United Kingdom, [6]Baylor Coll. of Med., Houston, TX, [7]Univ. Hosp. Bristol NHS Fndn. Trust, Bristol, United Kingdom, [8]King Faisal Specialist Hosp, Riyadh, Saudi Arabia, [9]Birmingham Womenâ□□s and Childrenâ□□s Hosp., Birmingham, United Kingdom, [10]Univ. of Louvain, Brussels, Belgium, [11]Inst. de Pathologie et de GÃ©nÃ©tique, Gosselies, Belgium, [12]Cincinnati Childrenâ□□s Hosp. Med. Ctr. and Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [13]GeneDx, Inc., Cincinnati, OH, [14]UCL Queen Square Inst. of Neurology, London, United Kingdom

Disclosure Block: **A. Verstraeten:** None.

Importin 8, encoded by *IPO8*, is an ubiquitously expressed member of the importin-Î² protein family that translocates cargo molecules such as proteins, RNAs and ribonucleoprotein complexes into the nucleus in a RanGTP-dependent manner. Current knowledge of the cargoes of importin 8 is limited, but TGF-Î² signaling components such as SMAD1-4 have been suggested to be amongst them. Here, we report that bi-allelic loss-of-function variants in *IPO8* cause a syndromic form of thoracic aortic aneurysm (TAA) with clinical overlap with Loeys-Dietz and Shprintzen-Goldberg syndrome. Seven individuals from six unrelated families showed a consistent phenotype with early-onset TAA, motor developmental delay, connective tissue findings and craniofacial dysmorphic features. A C57Bl/6N *Ipo8* knock-out mouse model recapitulates TAA development from 8-12 weeks onwards in both sexes, but most prominently shows ascending aorta dilatation with a propensity for dissection in males. Compliance assays suggest augmented passive stiffness of the ascending aorta in male *Ipo8*$^{-/-}$ mice throughout life. Immunohistological investigation of mutant aortic walls reveals elastic fiber disorganization and fragmentation along with a signature of increased TGF-Î² signaling, as evidenced by nuclear pSmad2 accumulation. RT-qPCR assays of the aortic wall in male *Ipo8*$^{-/-}$ mice demonstrate decreased *Smad6/7* and increased *Mmp2* and *Ccn2 (Ctgf)* expression, reinforcing a role for dysregulation of the TGF-Î² signaling pathway in TAA development. As importin 8 is the most downstream TGF-Î²-related effector implicated in TAA pathogenesis so far, it offers opportunities for future mechanistic studies and represents a candidate drug target for TAA.

PrgmNr 2165 - A novel *NOTCH3* terminal exon variant causes the rare lateral meningocele syndrome in an Asian child

View session detail

Author Block: **W. Ong**[1], M. Syuhebullah[1], D. Gunaseelan[1], N. Mohammad[2], W. T. Keng[1]; [1]Dept. of Genetics, Hosp. Kuala Lumpur, Kuala Lumpur, Malaysia, [2]Dept. of Radiology, Hosp. Sultan Ismail, Johor Bahru, Johor, Malaysia
Disclosure Block: **W. Ong:** None.
Lateral meningocele syndrome (LMS), or Lehman syndrome, is an exceedingly rare disorder characterized by multiple lateral spinal meningoceles with distinctive craniofacial features, hypotonia, joint hyperextensibility, developmental delay and variable skeletal, cardiac, and urogenital anomalies. It is caused by heterozygous pathogenic variants in the terminal exon 33 of *NOTCH3* gene, resulting in a truncated NOTCH3 intracellular domain protein with prolongation of its signalling effects in the highly conserved NOTCH signalling pathway. Most cases are de novo.
We present a 2.5-year-old girl, first child of a non-consanguineous Indonesian couple. She was a term baby with no antenatal nor postnatal complications. She had no feeding difficulties but her growth faltered postnatally and she was developmentally delayed particularly in gross motor and speech. Clinically she had a tall forehead, arched eyebrows, shallow supraorbital ridges with proptosis, midface hypoplasia, slightly upturned nose, thin and tented upper lip, full lower lip, micrognathia, low-set ears, pre-auricular pit, long halluces and hypotonia. There were no concerns with her hearing and vision. Chromosomal microarray, metabolic and biochemical indices and ultrasound of her urinary system were normal. WES identified a heterozygous variant, c.6705_6730del (p.Phe2235Leufs*10) in exon 33 of *NOTCH3* gene. This was classified a VUS as it was unreported in both disease and population databases, but multiple *in-silico* analyses suggest it disease-causing, and clinical features concurred with our patient. Following this result, MRI brain and spine revealed low-lying cerebellar tonsils, dural ectasia with multiple lateral meningoceles at thoracolumbosacral levels causing left posterolateral compression of the spinal cord at T11-L3 level, and mild posterior scalloping of the involved lumbar segments. Examination of parents were normal.
There had been fewer than 25 reported cases of LMS to our knowledge and our patient shared many of the clinical and radiological features previously described. This case also illustrated the power of reverse phenotyping in consolidating the clinical diagnosis, and in directing more precisely genetic counselling, long-term care and surveillance. This is especially important as lateral meningoceles can result in neurologic complications such as Chiari 1 malformation, neuropathy, paraparesis and bladder dysfunction. Other recommendations include cardiovascular monitoring for cardiac malformations and aortic dilatation, hearing assessment for mixed/conductive hearing loss, and ongoing neurodevelopmental and growth assessment.

PrgmNr 2166 - A novel variant in the *WNT10A* gene in a consanguineous Malian family with Odonto-Onycho-Dermal Dysplasia syndrome: First sub-Saharan African case

View session detail

Author Block: **A. YalcouyÃ©**[1,2], L. SidibÃ©[3], S. Diarra[1,4], L. CissÃ©[5], A. Maiga[1], K. Fischbeck[6], C. Guinto[1,5], G. LandourÃ©[7,6,5]; [1]FacultÃ© de MÃ©decine et dâ□□Odontostomatologie, USTTB, Mali, Bamako, Mali, [2]Univ. of cape Town, Cape Town, South Africa, [3]Service de PÃ©diatrie, CHU Gabriel TourÃ©, Bamako, Mali, Bamako, Mali, [4]Neurogenetics Branch, NINDS, NIH, Bethesda, MD, Bethesda, MD, [5]Service de Neurologie, CHU du Point Â« G Â», Bamako, Mali, Bamako, Mali, [6]Neurogenetics Branch, NINDS, NIH, Bethesda, MD, Bethesda, CA, [7]FacultÃ© de MÃ©decine et dâ□□Odontostomatologie, USTTB, Mali, Mali, Mali

Disclosure Block: **A. YalcouyÃ©:** None.

Introduction: Odonto-onycho-dermal dysplasia (OODD) is a rare autosomal recessive form of ectodermal dysplasia including severe oligodontia, nail dystrophy, palmoplantar hyperkeratosis, and hyperhidrosis. To date, no genetically confirmed case has been reported in Sub-Saharan Africa. Aim: To clinically characterize a Malian family with ectodermal dysplasia features and identify the underlying genetic defect.Methods: Patients and their relatives were ascertained upon a pediatrician initiated-referral. They were enrolled after giving written consent, and clinical data were obtained from medical records. DNA was extracted from peripheral blood via standard procedures for genetic analysis. Initially, the EDA, EDAR, EDARADD genes were tested in the proband. Then we sequenced the entire coding region of the WNT10A gene. PCR was performed with specific primers designed in our laboratory. Mutation screening was performed by a visually inspected for homozygous base changes and compared to the corresponding wild type sequences using the â□□BLASTâ□□. Once a nucleotide change was found, SNP databases were searched to determine its frequency. Pathogenicity was confirmed by in silico analysis using PolyPhen, SIFT and ClinVar. Results: We investigated a large consanguineous family of Fulani ethnicity. Six individuals of 11 to 28 years old were found to be affected. The disease onset was variable from birth to 15 years old with teeth agenesis and palmoplantar hyperkeratosis that worsened gradually. The clinical examination found a missing primary tooth ranging from 20 to 23 teeth, and palmoplantar hyperkeratosis itching with ulcerations in all the six patients. Sequencing of common ectodermal dysplasia genes (EDA, EDAR, and EDARADD) was negative. However, testing of the WNT10A gene identified a homozygous missense variant located at position c.1199G>A (p.Cys400Thr). This variant was not found in several SNP databases including in gnomAD/ExAC, dbSNP and 1000genomes, and is reported as pathogenic by several in silico prediction tools. Comments: We report here a novel homozygous missense variant in the WNT10A gene causing OODD syndrome. To the best of our knowledge, this is the first report of genetically confirmed OODD case with WNT10 mutation in the sub-Saharan African population, expanding its genetic epidemiology. Keywords: OODD, WNT10, novel variant, Mali, Africa

PrgmNr 2167 - A Palestinian patient with Wiedemann-Rautenstrauch like syndrome: expanding the phenotypic spectrum of PYCR1 mutations

View session detail

Author Block: **M. Ghannam**; Makassed Hosp., Jerusalem, Palestinian Territory
Disclosure Block: **M. Ghannam:** None.
Wiedemann-Rautenstrauch syndrome (WDRTS) is a rare autosomal recessive neonatal progeroid disorder characterized by intrauterine growth retardation, failure to thrive, progeroid appearance, decreased subcutaneous fat, hypotrichosis, hypotonia, and variable mental impairment. WDRTS is caused by biallelic variant in the POLR3A gene.
We report a female baby born for a consanguineous couple from Palestine. She presented with a relatively large head, distinctive facial features, progeroid appearance, and thin translucent skin with reduced subcutaneous fat. Trans-fontanel ultrasound showed agenesis of corpus callosum. Echocardiogram showed small patent ductus arteriosus, moderate patent foramen oval, and trivial peripheral pulmonary stenosis. Ophthalmic exam, skeletal survey, renal ultrasound, and karyotype were normal.
Whole exome sequencing revealed the presence of a homozygous pathogenic variant in *PYCR1* gene (c.616G>T)(p.Gly206Trp)). Recessive mutations in *PYCR1* were initially associated with Cutis Laxa Type IIB with progeroid features. However, further descriptions of patients bearing biallelic PYCR1 mutations revealed a remarkable clinical heterogeneity and phenotypic continuum ranging from isolated wrinkly skin syndromes, gerodermia osteo-dysplastica, De Barsy syndrome to severe progeria syndromes. The identified variant in our patient was reported before in a Palestinian child diagnosed with wrinkly skin syndrome and in another child form Qatar diagnosed with Geroderma osteodysplasticum. This highlight the phenotypic diversity that is associated with *PYCR1* gene even in patients with same genetic variant.
Clinical features in our patient were overlapping with that of WDRTS. Patient initially presented with upslanted palpebral fissures, thin translucent skin and reduced subcutaneous fat rather than wrinkled saggy skin appearance. However, patient reassessment at age of 40 days showed some lax and wrinkled skin with normal slanting palpebral fissures.
This report illustrates the wide spectrum of phenotypic features that could be associated with PYCR1 gene by showing that they can resemble WDRTS especially in the early neonatal period.

PrgmNr 2168 - Acromesomelic skeletal dysplasia with severe short stature due to a biallelic *KIF24* variant

View session detail

Author Block: **N. u. Ain**[1], T. Ishaq[2], O. Makitie[3,1], S. Naz[4]; [1]Karolinska, Inst.t, Stockholm, Sweden, [2]Univ. of the Punjab, Lahore, Pakistan, Lahore, Pakistan, [3]Univ. of Helsinki, Helsinki, Finland, [4]Univ. of the Punjab, Lahore, Pakistan

Disclosure Block: **N.U. Ain:** None.

Several types of skeletal dysplasia have been described and molecularly characterized. Studies have identified skeletal dysplasia to occur due to impaired ciliogenesis in some cases. We investigated a Pakistani consanguineous family to underpin the genetic cause of skeletal dysplasia, similar in phenotype to acromesomelic dysplasia. Clinical and radiographic investigations were performed for two affected individuals of the family. Whole genome sequencing (WGS) was completed using DNA from one affected family member. Data obtained after WGS was analyzed and filtered to identify the possible disease causing variant. Sanger sequencing was performed to confirm the segregation of the identified variant with the disease phenotype and to determine its allele frequency in controls. Affected individuals in the family had severe short stature (KIF24, which segregated with the disease phenotype in the family. The variant was absent from public databases as well as DNA of 200 ethnically matched controls. *KIF24* encodes a 1368 amino acids protein belonging to the kinesin family. KIF24 has a role in ciliogenesis and microtubule polymerization. The identified variant affects an amino acid that is conserved not only in all vertebrate species, but also among different kinesin proteins. This study describes a novel form of skeletal dysplasia associated with biallelic variant in *KIF24*.

PrgmNr 2169 - Clinically actionable secondary findings in 390 whole genome sequence data from sub-Saharan African families with nonsyndromic orofacial clefts

View session detail

Author Block: **L. Gowans**[1], A. Oladayo[2], W. O. Awotoye[2], A. Alade[3], B. Tamara[2], T. Naicke[4], M. A. Eshete[5], W. L. Adeyemo[6], J. Bidinger[7], E. Zheng[2], O. Adamson[6], C. Adeleke[2], M. Li[2], S. Kayali[2], J. Olotu[8], P. A. Mossey[9], S. Obiri-Yeboah[1], T. H. Beaty[10], M. Taub[11], P. Donkor[1], M. L. Marazita[12], A. A. Adeyemo[13], J. C. Murray[14], A. Butali[15]; [1]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [2]Univ. of Iowa, Iowa City, IA, [3]Iowa, IA, [4]Univ. of KwaZulu-Natal, Durban, South Africa, [5]Addis Ababa Univ., Addis Ababa, Ethiopia, [6]Univ. of Lagos, Lagos, Nigeria, [7]Johns Hopkins Univ., Baltimore, MD, [8]Univ. of Port Harcourt, Port Harcourt, Nigeria, [9]Univ. of Dundee, Dundee, United Kingdom, [10]Johns Hopkins Univ, Sch PubHlth, Baltimore, MD, [11]Johns Hopkins, Baltimore, MD, [12]Univ Pittsburgh, Pittsburgh, PA, [13]NIH, Bethesda, MD, [14]U of Iowa, Iowa City, IA, [15]Univ Iowa, Iowa City, IA
Disclosure Block: **L. Gowans:** None.

In 2017, the American College of Medical Genetics and Genomics (ACMG) argued for the need to return secondary findings in genomic data on 59 genes that are associated with conditions that are clinically actionable. This list of 59 genes was updated further to 73 genes by ACMG in 2021. In this study, we carried out whole genome sequencing of 150 case-parent trios; only the probands had orofacial clefts. Participants were recruited from Ghana and Nigeria. Of the 150 families, a total of 130 families passed various quality control checks and were screened for secondary findings based on the 59 clinically actionable genes released by ACMG in 2017. In all, we observed 144 unique genetic variants in 44 out of the 59 genes released by ACMG in 2017. Of the 144 unique variants, about 97% (140) were variants of unknown significance (VUS) whereas about 3% (4) of the variants were classified as pathogenic or likely pathogenic. A heterozygous missense variant in *PRKAG2* gene (p.Glu183Lys) was classified as pathogenic, whereas three other heterozygous variants in *RYR1* (p.Arg2163Leu) and *LDLR* (p.Ala431Thr and p.Asn564Ser) were predicted to be likely pathogenic. The p.Asn564Ser variant in *LDLR* was observed in a mother in one of the families. The other three variants (p.Glu183Lys, p.Arg2163Leu and p.Ala431Thr) were observed in both mothers and probands in three different families. *PRKAG2, RYR1* and *LDLR* are associated with Wolff-Parkinson-White (WPW) Syndrome, Malignant hyperthermia and Familial hypercholesterolemia 1, respectively. These conditions are all clinically actionable. We will therefore follow-up on these patients and collaborate with clinicians to offer the necessary genetic counseling and other clinical care. In conclusion, we have shown that clinically actionable secondary findings were observed in 1.79% (7/390) of participants, which is slightly higher than about 1% reported for diverse ethnicities. Our results further corroborate the ACMGâ□□s position on the need to intentionally search clinical genomics sequence data for actionable secondary findings based on the ACMGâ□□s list of actionable genes.

PrgmNr 2170 - Congenital bilateral anophthalmia: A case report

View session detail

Author Block: **A. Kondo**[1], A. Hayashi[1], A. Tachibana[1], M. Morine[1], K. Hinokio[1], T. Kirino[1], T. Iwai[1], Y. Kuroki[1], N. Okamoto[2], K. Maeda[1]; [1]Shikoku Med. Ctr. for Children and Adults, Zentsuji, Japan, [2]Osaka Women's and Children's Hosp., Osaka, Japan

Disclosure Block: **A. Kondo:** None.

Anophthalmia means absence of ocular tissue in the orbit and it can be unilateral or bilateral. It may occur in isolation or as part of a syndromic condition. Association with other systemic malformations often involving cardiac, musculoskeletal and central nervous system. A few known pathogens that can cause anophthalmia are Toxoplasma, rubella, and certain strains of the influenza virus. Other known environmental conditions that have led to anophthalmia are maternal vitamin A deficiency, exposure to X-rays during gestation, solvent abuse, and exposure to thalidomide. Some chromosomal abnormalities are reported associated with this condition, but as for monogenic causes, only SOX2 has been identified as a major causative gene.A boy was bone at 37w4d, 2632g, Apg7/10 as normal vaginal delivery. He was referred to childrenâ□□s hospital because of blepharophimosis at first. Head CT showed very small eyeballs and Extraocular muscles and Rt optic nerve. General examination systemic work-up showed that he had hearing issues and cryptorchidism. His weight gain was poor and it was related to dysphagia. Until he got 5 y/o, parents were not interested in making diagnosis, but later at age 6 y/o, he had genetic analysis at different hospital and it revealed a mutation in SOX2-OT gene which is a causative gene of his condition. 6 years later, mother showed up to our clinic to ask the meaning of genetic test. At that time, he was 12 y/o and had some more symptoms, such as epilepsy, GH insufficiency, hypogonadism. Head MRI showed thin corpus callosum and thin brain stem. His development was delayed as suspected before, however he has been recognising his mother gradually and could turn over himself even he cannot hold a sitting position. He is now enjoying his school for the visually impaired person. His mother was happy with his development and having settled daily life with him. Her worry was the possibility of inheritance of this condition to his younger brotherâ□□s future children at last. Genomic test for making definite diagnosis is done as general clinical medicine these days. However, this patient and his family have not been interested in the genomic test so much consistently. They have decided to have genetic analysis because of strong recommendation from local paediatrician and the family did not understand well about the result. It is very informative for us to know genetic mutation and phenotype, still we have to remember that patient familyâ□□s wish is more about the way to find happy life of their children. Some patient family need more easy, relaxed â□□Genetic counsellingâ□□ easing their fear rather than informative lecture.

PrgmNr 2171 - C-terminal and N-terminal truncating mutations of the *MN1* gene lead to distinct developmental phenotypes

View session detail

Author Block: **H-Y. B. Chung**[1], C. C. Y. Mak[1], M. M. C. Chui[1], M. Lee[1], N. Vegas[2], A. E. Lin[3], P. A. Sanchez-Lara[4], H. M. Bombei[5], J. A. Bernat[5], L. Faivre[6], F. Tran-Mau-Them[7], I. Valenzuela Palafoll[8], K. A. Metcalfe[9], N. Canham[10], A. Sarkar[11], Y. A. Zarate[12], B. Callewaert[13], MN1 International Study Group, P. Kruszka[14], J. Amiel[2], D. Doherty[15], C. T. Gordon[2]; [1]Dept. of Paediatrics & Adolescent Med., LKS Faculty of Med., The Univ. of Hong Kong, Hong Kong, China, [2]Lab. of Embryology and Genetics of Human Malformations, Inst. Natl. de la SantÃ© et de la Recherche MÃ©dicale (INSERM) UMR 1163, Inst. Imagine, UniversitÃ© de Paris, 75015 Paris, France, [3]Med. Genetics, MassGen. Hosp. for Children, Boston, MA, [4]Dept. of Pediatrics, Cedars-Sinai Med. Ctr., Los Angeles, CA, [5]Div. of Med. Genetics and Genomics, Stead Family Dept. of Pediatrics, Univ. of Iowa Hosp., Iowa City, IA, [6]Ctr. de RÃ©fÃ©rence Anomalies du DÃ©veloppement et Syndromes Malformatifs, CHU Dijon Bourgogne, Dijon, France, [7]UF Innovation en diagnostic gÃ©nomique des maladies raresUF Innovation en diagnostic gÃ©nomique des maladies rares, Ctr. Hosp.ier Univ.ire de Dijon, Dijon, France, [8]Dept. of Clinical and Molecular Genetics, Univ. Hosp. Vall dÂ´Hebron and Med. Genetics Group, Valle Hebron Res. Inst., Barcelona, Spain, [9]Manchester Ctr. for Genomic Med., St Mary's Hosp., Manchester Univ. NHS Fndn. Trust, Hlth.Innovation Manchester, Manchester, United Kingdom, [10]Liverpool Ctr. for Genomic Med., Liverpool Women's Hosp., Crown Street, Liverpool, United Kingdom, [11]Dept. of Clinical Genetics, Nottingham Univ. Hosp. Natl. Hlth.Service Trust, Nottingham NG5 1PB, United Kingdom, [12]Section of Genetics and Metabolism, Univ. of Arkansas for Med. Sci., Little Rock, AR, [13]Ctr. for Med. Genetics Ghent, Ghent Univ. Hosp., Ghent 9000, Belgium, [14]GeneDx, Gaithersburg, MD, [15]Dept. of Pediatrics, Univ. of Washington, Seattle, WA

Disclosure Block: **H.B. Chung:** None.

MN1 C-Terminal Truncation (MCTT) syndrome is a rare autosomal dominant disorder characterized by intellectual disability, mid-face hypoplasia, severe expressive speech delay, and an atypical form of rhombencephalosynapsis (Mak et al. 2020 and GeneReviewsÂ®). The *MN1* gene is comprised of only two exons. Individuals with truncating variants at the C-terminal end of *MN1* (within exon 2 or the last 55bp of exon 1, and therefore escaping nonsense-mediated decay (NMD)) present with the distinct syndromic features of MCTT. Only seven individuals affected by N-terminal truncations (predicted to induce NMD) have been published so far, and have milder developmental phenotypes than MCTT patients (Mak et al. 2020, Shu et al. 2021 and Vegas et al. 2021). We performed a multicentre study to review an expanded case series (n=45) of 27 male and 18 female subjects (M:F ratio 1.5:1, mean age 12.9, range 2-44) from North America, Europe and Asia to characterize and compare the clinical and molecular spectrum of C-terminal (n=32) and N-terminal (n= 13) truncating mutations. Thirteen previously unreported patients (29%) are presented here. Review of 25 known and 7 unreported MCTT cases show that the characteristic dysmorphic features are consistently observed, which include midface hypoplasia, downslanting palpebral fissures, hypertelorism, exophthalmia, short upturned nose, and dysplastic low-set ears. Three recurrent variants are found (c.3778G>T, c.3883C>T and c.3903G>A), with the c.3883C>T variant occurring in 9 of 32 individuals (28%). We also report the oldest MCTT patient known to date, a 44-year-old male diagnosed at the age of 40. All patients have some degree of intellectual disability (100%), motor delay (100%) and speech delay (100%). For expressive speech delay, 33% of individuals with MCTT rely on non-verbal communication only, with the remaining expressing first words at the mean age of 4.03 (range 2-6.75 years). Remarkably, of the 13 individuals with N-terminal variants, one had typical facial dysmorphic features of MCTT. Cleft palate (33% vs 7%) and conductive hearing loss (82% vs 35%) are more commonly observed in association with N-terminal vs C-terminal truncations. Fewer individuals affected by N-

terminal truncations have intellectual disability (17% vs 100%) and motor delay (33% vs 100%). Despite this, 73% have some degree of expressive speech delay with first words at the mean age of age 2 (range 1.3-3 years). Globally our study supports the previous observation that patients with N-terminal truncations have milder neurodevelopmental outcomes than MCTT patients.

PrgmNr 2172 - Delineation of the phenotype and genotype in ten individuals with de novo variants in ZBTB18

View session detail

Author Block: **R. Thomas**[1], S. Joss[2], N. Lahiri[3], T. Homfray[1], J. Vogt[4], A. Barnicoat[5], A. Male[6], J. Cobben[7], C. J. Shaw-Smith[8], P. Constantinou[2], T. Linton-Willoughby[1], R. Laskowski[9], W. D. Jones[10]; [1]St George's Univ. Hosp., London, United Kingdom, [2]Queen Elizabeth Univ. Hosp., Glasgow, United Kingdom, [3]St George S Univ. of London, London, United Kingdom, [4]Birmingham Women's NHS Fndn. Trust, Birmingham, United Kingdom, [5]Great Ormond Street Hosp. for Children NHS foundation trust, London, United Kingdom, [6]Great Ormond Street Hosp. for Children NHS Fndn. Trust, London, United Kingdom, [7]North West Thames Regional Genetics Service, London North West Univ. Hlth.care NHS Trust, London, United Kingdom, [8]ROYAL DEVON AND EXETER NHS FOUNDATION TRUST, Exeter, United Kingdom, [9]European Bioinformatics Inst. (EMBL-EBI), European Molecular Biology Lab., Cambridge, United Kingdom, [10]Great Ormond Street Hosp., London, United Kingdom, United Kingdom
Disclosure Block: **R. Thomas:** None.
Multiple sources have hypothesised that pathogenic variation within the gene ZBTB18 contributes to 1q43/44 microdeletion syndrome. This condition is characterised by intellectual disability, seizures, corpus callosum anomalies along with shared dysmorphic features. Previously a number of individuals have been identified with variants in ZBTB18 in the literature, although phenotypic information is sparse and inconsistent. We present the clinical details of 10 individuals with de novo variants in ZBTB18, classified as pathogenic or likely pathogenic, identified from the Deciphering Developmental Disorders study. We report previously unreported features such as sleep disturbance and hypertrichosis. We present facial features, growth parameters, interrogation of developmental milestones along with details of neuroimaging. Our work also involves detailed examination of each variant identified in conjunction with analysis from protein modelling experiments. This work strengthens the case for the inclusion of ZBTB18 on gene panels for intellectual disability. We hope that by sharing further phenotypic information this work will aid clinicians in considering ZBTB18 when formulating a differential diagnosis or interpreting variation within this gene.

PrgmNr 2173 - Expanding the phenotype of cerebello-facio-dental syndrome in female with a novel pathogenic variant

View session detail

Author Block: **H. Yoshihashi**[1], K. Nagahara[2], H. Futagawa[1], S. Ito[3], M. Honda[4], M. Yamada[5], H. Suzuki[5], T. Uehara[6], T. Takenouchi[5], K. Kosaki[5]; [1]Dept. of Clinical Genetics, Tokyo Metropolitan Childrenâ□□s Med. Ctr., Tokyo, Japan, [2]Dept. of Pediatrics, Showa Univ. Sch. of Med., Tokyo, Japan, [3]Dept. of Nursing, Tokyo Metropolitan Childrenâ□□s Med. Ctr., Tokyo, Japan, [4]Dept. of Clinical Res., Tokyo Metropolitan Childrenâ□□s Med. Ctr., Tokyo, Japan, [5]Ctr. for Med. Genetics, Keio Univ., Tokyo, Japan, [6]Aichi Dev.al Disability Ctr. Central Hosp.,Inst. for Dev.al Res., Aichi-ken, Japan
Disclosure Block: **H. Yoshihashi:** None.
Cerebello-facio-dental syndrome (MIM#616202:CFDS) is an autosomal recessive condition characterized by intellectual disability, microcephaly, cerebellar hypoplasia, dysmorphic features and short stature. In CFDS, six different pathogenic variants in BRF1 gene (5 missense and 1 nonsense variants) have been reported. Herein we additionally present a female with CFDS, expanding the phenotype and involving a novel pathogenic variant. The patient is 5-year-old girl of healthy and non-consanguineous parents. Birth weight was 2807g. She was born at 40 weeks of gestations after uneventful pregnancy and needed to have operations for congenital heart malformations(DORV,VSD)and laryngomalacia. In infancy, she repeatedly showed severe conditions caused by respiratory tract infection, resulting in the need for hospitalization and IVIGs. At age of 2 years, short stature, failure to thrive, microcephaly: OFC 42.5cm at 33months(-4.5SD) and severe global developmental delay have emerged. Facial dysmorphism includes sparse eyebrows, almond-shaped eyes, down-turned corners of mouth and prominent upper incisors. Kyphosis, hypothyroidism and hypermetropic astigmatism were observed since 3 years of age. Brain MRI showed a cerebellar hypoplasia. Postnatal marked growth retardation was also prominent at age of 5 years; height 84.6cm(-5.4SD), weight 10.3kg(-3.0SD) . G-banded chromosomal analysis revealed normal female karyotype. SNP-array detected no causative copy number variant. Whole exome sequencing identified compound heterozygous variants in BRF1 gene; p.(Thr259Met) previously reported and p.(Ser293Leufs*15), segregated in the parents. Thus far, ten biallelic carriers affected with CFDS have been reported. We detected additional two findings not reported to date, including thyroid dysfunction and repetitively severe infection in infancy. The latter could be a potential life-threatening manifestation in infancy, because there is one deceased infant at 21 months of age due to respiratory complications in a previous report. As noted above, both conditions should be screened in diagnosed infants with CFDS. The further description of patients with CFDS is essential to better delineate the phenotypic and genotypic spectrum.

PrgmNr 2174 - Long-read whole genome sequencing identified a partial *MBD5* deletion in an exome-negative patient with neurodevelopmental disorder

View session detail

Author Block: **S. Ohori**, N. Matsumoto; Yokohama City Univ. Graduate Sch. of Med., Yokohama, Kanagawa, Japan
Disclosure Block: **S. Ohori:** None.
Whole exome sequencing (WES) can detect not only single nucleotide variants in causal genes, but also pathogenic copy number variations using several methods. However, there may be overlooked pathogenic variations in the out of target genome regions of WES analysis (e.g. promoters), leaving many patients undiagnosed. Whole genome sequencing (WGS) can potentially analyze such regions. We applied long-read nanopore WGS and our recently developed analysis pipeline â□□dnarrangeâ□□ to a patient who was undiagnosed by trio-based WES analysis, and identified a heterozygous 97 kb deletion partially involving 5â□□- untranslated exons of *MBD5*, which was outside the WES target regions. The phenotype of the patient, a 32-year-old male, was consistent with haploinsufficiency of *MBD5*. The deletion breakpoint PCR analysis in the patient and his parents showed that the deletion of *MBD5* occurred de novo. The transcript level of *MBD5* in the patientâ□□s lymphoblastoid cells was reduced. American College of Medical Genetics and Genomics (ACMG) guideline for the interpretation of copy-number variants suggests that this de novo deletion partially involving 5â□□ untranslated region of *MBD5* was likely pathogenic (0.90) for the patient based on evidences of 1A (0), 2C-2 (0.45), 3A (0) and 5A (0.45). Taken together, we considered the 29-kb partial *MBD5* deletion as disease-causing SV. Furthermore, we found other rare structural variations (SVs) in this patient, i.e., a large inversion and a retrotransposon insertion, which were not seen in 33 controls. Although we considered that they are benign SVs, this finding suggests that our pipeline using long-read WGS is useful for investigating various types of potentially pathogenic SVs. In conclusion, we identified a 97-kb deletion which causes haploinsufficiency of *MBD5* in a patient with neurodevelopmental disorder, demonstrating that long-read WGS is a powerful technique to discover pathogenic SVs.
Acknowledgment: We would like to thank Drs. Rie S. Tsuburaya, Masako Kinoshita, Etsuko Miyagi, Takeshi Mizuguchi, Satomi Mitsuhashi, Martin C Frith for their participation to this study.

PrgmNr 2175 - Loss-of-Function mutations of *USP9X* as the differential diagnosis for CHARGE syndrome

View session detail

Author Block: **H. Futagawa**[1], S. Ito[1], M. Honda[1], H. Suzuki[2], M. Yamada[3], T. Uehara[4], T. Takenouchi[5], K. Kosaki[2], H. Yoshihashi[6]; [1]Tokyo Metropolitan Childrenâ□□s Med. Ctr., Fuchu, Japan, [2]Keio Univ. Sch. of Med., Tokyo, Japan, [3]Ctr. for Med. Genetics, Keio Univ., Tokyo, Shinjuku-ku, Japan, [4]Aichi Dev.al Disability Ctr. Central Hosp.,Inst. for Dev.al Res., Aichi-ken, Japan, [5]Keio Univ, Tokyo, Japan, [6]Tolyo Metropolitan Children's Hos, Fuchu-shi, Japan

Disclosure Block: **H. Futagawa:** None.

Loss-of-function mutations in *USP9X*(Xp11.4) is known to cause Mental retardation, X-linked 99, syndromic, female-restricted (MIM#300968:MRXS99F). Clinical features include developmental delay, distinctive facial features, short stature and multiple congenital anomalies, for example choanal atresia, heart anomalies and anal anomalies. These features may overlap with CAHRGE syndrome (CS). Here, we report additionally two Japanese cases of MRXS99F and discussed the differentiation of MRXS99F and CS. 3-year-old girl. At the second trimester, she diagnosed cystic hygroma and conventional chromosomal analysis revealed normal female karyotype. After birth, she diagnosed congenital heart anomalies, anal atresia, unilateral hydronephrosis, lower limb asymmetry and hypopigmentation along Blaschko line. Her dysmorphic facial features included prominent forehead, long face, facial hemangioma, short palpebral fissures, symmetrically bilateral ear malformations, long nose, smooth and long philtrum. Now, her height and body weight were -1.4SD and -1.8SD respectively and she had global developmental delay. WES (IRUD) revealed de novo pathogenic variant in *USP9X* (NM_00103950.2:c.1153C>T p.Arg385*). 9-year-old girl. Her mother had keratoconus. After birth, she diagnosed anal atresia, unilateral choanal atresia and hearing loss, and she was suspected for CS. Now, her height and body weight were -3.6SD and -2.5SD respectively and she had global developmental delay. Her dysmorphic facial features included relative macrocephaly, square face, prominent forehead, hypertelorism, symmetrically bilateral ear anomalies and long philtrum. WES revealed de novo missense variant in *USP9X* (NM_00103950.2:c.3148G>A p.Asp1050Asn). Patients with MRXS99F have some clinical features overlap with CS. Our case 2 was suspected for CS. Present cases did not have coloboma, cranial nerve abnormalities and asymmetrical ear anomalies, which help to define the diagnosis of CS. These features rarely have reported with MRXS99F. Detailed physical assessment would make it possible to distinguish these two syndromes, and atypical CS may be the clue for the diagnosis of MRXS99F. Reviewing the literature, prominent forehead, low nasal bridge, prominent nose, long philtrum and ear anomalies were considered as distinctive facial features. Prominent forehead, square face and long philtrum might be the characteristic facial manifestations and MRXS99F would be categorized as a recognizable syndrome. The female case of atypical CHARGE syndrome with facial dysmorphism may be suggestive features of MRXS99.

PrgmNr 2176 - Objective evaluation of dysmorphism by automated analysis of facial photographs

View session detail

Author Block: **G. Mubungu**[1], P. Makay[1], N. Fleischer[2], A. Verloes[3], L. Mutesa[4], M. Roelants[5], P. Lukusa[1], K. Devriendt[5], A. Lumaka Zola[6]; [1]UniversitÃ© de Kinshasa, Kinshasa, Congo, Democratic Republic of the, [2]FDNA Inc., Boston, MA, [3]Robert DEBRE Univ. Hosp., Paris, France, [4]Univ. of Rwanda, Kigali, Rwanda, [5]KU Leuven, Leuven, Belgium, [6]UniversitÃ© de LiÃ¨ge, LEUVEN, Belgium
Disclosure Block: **G. Mubungu:** None.
The evaluation of â□□dysmorphismâ□□ remains challenging. The presence of three or more minor anomalies is often used as clinical criterion. However, this requires expertise in the evaluation of facial features. The fields of dysmorphology and syndromology are rapidly advancing, triggered by developments in testing technologies such as next generation sequencing, as well as well as automated sorting of syndromes. There is a need for tools that offer a more objective and faster recognition of dysmorphism and syndromes. Face2Gene (F2G) is a widely adopted tool for recognition of syndromes from 2D facial photographs.In the present study, we explored how this tool could aid in the recognition of facial dysmorphism and how it could react to factors such as age, sex and ethnic background. Different cohorts were studied including unselected Congolese newborns, Congolese children with intellectual disability, children with Down syndrome (DS) from DR Congo and Rwanda (African ethnicity), from Belgium (Caucasian) and Guadeloupe (mixed ethnicity), and a cohort of healthy adult Congolese volunteers.We used F2G to extract facial features from facial photographs, calculate dysmorphism scores and study the effect of ethnic origin, age and gender in different cohorts.We observed that F2G overestimated the incidence of individual minor facial features in the cohort of Congolese newborns. F2G detected facial dysmorphism, defined as the simultaneous presence of three or more minor facial anomalies, with a sensitivity of 37.5% and a specificity between 94-98%. This suggest that F2G performs better in a holistic approach rather than feature based approach in African individuals.In addition, F2G was able to clearly distinguish Congolese children aged above 15 years from those between 10-14 years based on their facial photographs (AUC=0.874). F2G also distinguished Congolese boys versus girls only from the age of 25 years (AUC=0.998). We concluded that age and gender play a significant role in baseline morphology and in dysmorphism after puberty.It was not possible to separate unselected Congolese newborns based on the geographical province of parents within DRC. Interestingly, a clear distinction was made between children with DS from different countries. The African (DR Congo and Ruanda) DS patients were very distinct from Caucasian DS patients from Belgium (AUC=1.000 and AUC=1.000) within the same range of age. Moreover, mixed ethnicity DS patients from Guadeloupe were clearly distinct from Belgian patients (AUC=1.000) but closer to Congolese DS patients (AUC=0.741). This suggests that ancestral genetic background influences the phenotypic expression of DS.

PrgmNr 2177 - Say-Barber-Biesecker-Young-Simpson syndrome (SBBYSS) and the importance of Whole Exome Sequencing in finding the accurate diagnostic. A case report

View session detail

Author Block: **D. Sabau**[1], L. C. Bohiltea[2], V. E. Radoi[2,1]; [1]Personal Genetics, Bucharest, Romania, [2]Carol Davila Univ. of Med. and Pharmacy, Bucharest, Romania
Disclosure Block: **D. Sabau:** None.
Introduction: Say-Barber-Biesecker-Young-Simpson variant of Ohdo syndrome (SBBYSS) is a genetic, autosomal dominant syndrome caused by *KAT6B* mutations. *KAT6B* - related disorders have variable expression and are characterized by developmental delay, hypotonia, dysmorphic features, congenital cardiac defects, genital abnormalities, skeletal abnormalities, dental anomalies.
Material and methods: A 3-year-old Caucasian boy was referred to the Department of Clinical Genetics because of developmental delay and dysmorphic features. The boy is the first child of healthy unrelated parents. He was born at term, with a gestational age of 40 weeks and one day and his birth weight was 3,425 g and length was 52 cm. The clinical evaluation revealed developmental delay, dismorphic features: hypertelorism, epicanthus, blepharophimosis, small and dystrophic teeth, long toes, atrial septal defect, and generalized hypotonia, bilateral cryptorchidism. Metabolic screening, array CGH and WES were performed for the patient.
Results: Metabolic screening and array CGH (8x60k Agilent) had both normal results. The next step in diagnosis algorithm was to perform WES analysis. The result of WES was an atypical *KAT6B* mutation, a de novo synonymous variant, located in exon 16 (c.3147G>A, p.(Pro1049Pro)). This mutation was previously identified in three unrelated patients. This exonic mutation was predicted in silico to cause protein truncation through aberrant splicing.
Conclusion: Our medical team decided to perform WES instead of *KAT6B* mutation screening, and recommends this method for further investigation of *KAT6B* - related phenotypes. There is no doubt that WES has increased the success rate when trying to identify the genetic cause of rare Mendelian disorders.

PrgmNr 2178 - The first East Asian adult case with rare *NEPRO*related skeletal dysplasia caused by a novel missense homozygous variant in *NEPRO*

View session detail

Author Block: **M. Minatogawa**[1,2], Y. Tsukahara[3], A. Sawaumi[4], T. Kosho[1,2,5]; [1]Dept. of Med. Genetics, Shinshu Univ. Sch. of Med., Matsumoto, Japan, [2]Ctr. for Med. Genetics, Shinshu Univ. Hosp., Matsumoto, Japan, [3]Dept. of Radiology, Shinshu Univ. Sch. of Med., Matsumoto, Japan, [4]Dept. of Orthopedics, Azumino Red Cross Hosp., Azumino, Japan, [5]Res. Ctr. for Supports to Advanced Sci., Matsumoto, Japan

Disclosure Block: **M. Minatogawa:** None.

Background: Biallelic pathogenic variants in *NEPRO* is known to cause one of the skeletal dysplasia characterized by severe short stature, brachydactyly, skin laxity, joint hypermobility, and radiographic skeletal abnormalities. *NEPRO* related skeletal dysplasia (NEPSD) was first reported by Shaheen et al. in 2016 and is an ultrarare disease with only 5 children with NEPSD reported so far. Here we report the first individual of NEPSD in a Japanese adult female with a novel missense homozygous variant in *NEPRO*. **Case report:** The propositus is a 44-year-old woman with clinical features including short stature of 130cm (â□□5.7SD), brachydactyly of the digits and toes, hypoplastic nails, pectus excavatum, skin laxity, joint hypermobility, and scoliosis. She also presented with congenital hip dislocation and acetabular dysplasia requiring surgery, epilepsy controlled by pharmacological treatment, bilateral congenital cataracts, and intentional myoclonus. She had constant joint pain due to severe joint hypermobility, and she gradually became difficult to walk. As a result, she has lived in a wheelchair since late childhood and has always used painkillers. Whole exome sequencing (WES) revealed a novel missense variant homozygously in *NEPRO* (NM_015412:exon4:c.442C>G:p.Arg148Gly). **Conclusion:** This is the first detailed description of the clinical course of NEPSD from childhood to adulthood, which will expand the knowledge of the natural history of this syndrome. The present case showed short stature, brachydactyly, skin laxity, joint hypermobility, acetabular hypoplasia, and scoliosis, which are those reported in previous NEPSD cases. We have also found several findings (congenital cataracts, seizures, and intentional myoclonus) in the present case that have been unreported in NEPSD so far. Although no variants in the genes responsible for cataracts and myoclonus were detected in WES, it remains to be clarified whether cataracts and myoclonus were associated with NEPSD. Further functional analysis and accumulation of cases would clarify these issues.

PrgmNr 2179 - The genotypic and phenotypic spectrum of pycnodysostosis in Saudi Arabia: Novel variants and clinical findings

View session detail

Author Block: **A. Mushiba**[1], E. A. Faqeih[2], M. Saleh[1], K. Ramzan[3], F. Imtiaz[4], M. AlOwain[5], A. Alhashem[1], A. F. Alswaid[6]; [1]King Fahad Med. City, Riyadh, Saudi Arabia, [2]King Fahad Med City, Children Specialist Hosp., Riyadh, Saudi Arabia, [3]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Riyadh, Saudi Arabia, [4]KFSH, Riyadh, Saudi Arabia, [5]King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia, [6]King Abdulaziz Med City, Riyadh, Saudi Arabia

Disclosure Block: **A. Mushiba:** None.

Pycnodysostosis is characterized by short stature, osteosclerosis, acro-osteolysis, increased tendency of fractures, and distinctive dysmorphic features. It is a rare autosomal recessive disease caused by biallelic CTSK mutations. The clinical details of 18 patients from Saudi Arabia were reviewed. Short stature, osteopetrosis, acroosteolysis, and distinctive facial dysmorphism were documented in all cases. Our results highlight the significant complications associated with this disease. The large anterior fontanelle is one of the cardinal signs of this disease; however, half of our patients had small fontanelles and a quarter had craniosynostosis, which caused optic nerve compression. Sleep apnea was of the major complications in three patients. Bone fracture can be a presenting symptom, and in our patients it mainly occurred after the age of 3 years. Bone marrow suppression was seen in a single patient of our cohort who was misdiagnosed initially with malignant osteopetrosis. In this study, we also describe two novel (c.5G > A [p.Trp2Ter], c.538G > A [p.Gly180Ser]) and two reported (c.244-29 A > G, c.830C > T [p.Ala277Val]) CTSK mutations. Our results indicate that the recurrent intronic variant, c.244-29 A > G is likely to be a founder mutation, as it was found in 78% (14/18 patients) of our cohort belonging to the same tribe.

PrgmNr 2182 - A novel variant in the *GNE* gene in a Malian with distal muscle weakness

View session detail

Author Block: **A. Maiga**[1], S. Bamba[1], S. Diarra[2,1], S. Diallo[3], S. Diallo[1,3], T. Coulibaly[1,4], A. YalcouyÃ©[1,5], L. CissÃ©[1,6], B. Maiga[7], G. LandourÃ©[1,4], C. Guinto[1,4]; [1]Univ. of Sci., Techniques and Technologies of Bamako, Bamako, Mali, [2]NIH, ROCKVILLE, MD, [3]Ctr. Hosp.ier Univ.ire 'Gabriel TourÃ©', Bamako, Mali, [4]Ctr. Hosp.ier Univ.ire 'Point G', Bamako, Mali, [5]Univ. of CapeTown, CapeTown, South Africa, [6]HÃ´pital Namakoro Fomba, Segou, Mali, [7]HÃ´pital SominÃ© Dolo, Mopti, Mali
Disclosure Block: **A. Maiga:** None.
Introduction: GNE myopathy is an ultra-rare autosomal recessive muscle disease with an estimated prevalence of 1 to 9:1,000,000. It is characterized by progressive distal muscle weakness and ultimately leads to wheelchair binding. The disease is caused by a mutation in the *GNE* gene that encodes the enzyme sialic acid epimerase, which is responsible for the last step of sialic acid biosynthesis. Although close to 200 variants were reported worldwide in different populations, cases in sub-Saharan Africa are scarce. **Aims**: To characterize both clinically and genetically patients with distal muscle weakness in the Malian population **Methods**: Patients with distal muscle weakness and their relatives were evaluated, after giving written consent, by a multidisciplinary clinical team including neurologists and cardiologists. Electromyography was performed and peripheral was collected for blood chemistries and to extract DNA for genetic testing. **Result**: A 19-year-old man from a consanguineous family was seen for progressive walking difficulty. The disease started at age 17 with a walking difficulty that worsened over time and has led to frequent falls. Clinical examination was consistent with distal muscle weakness and atrophy, reduced reflexes, and distal sensory loss. EMG has shown reduced distal motor and sensory amplitudes. These clinical and laboratory findings were in favor of distal neuropathy. However, CMT gene panel testing was negative. But WES revealed a novel pathogenic missense variant in the *GNE* gene (c.1841G>T [p.S614I]). This mutation segregates with the disease status in the family. **Conclusion**: This is likely going to be the first *GNE*-related myopathy diagnosed in the continent but the second reported in the sub-Saharan African population; suggesting that genetic diseases, in general, are underexplored in the African population, and that a better clinical assessment and access to diagnostic tools could uncover many others in this part of the World. Keywords: *GNE* gene, novel variant, Mali, Africa.

PrgmNr 2183 - A rare double diagnosis identified via exome sequencing in a patient with complex cerebellar ataxia

View session detail

Author Block: **Q. THOMAS**[1,2,3], F. Tran Mau-Them[4], P. Garret[2], C. Racine[1], V. Leguy[5], C. Philippe[2,4], A. Charlot[6], M. Sauvant[3], A. Jacquin-Piques[7], K. Guillen[8], Y. BÃ©jot[3], L. Faivre[1], Y. Duffourd[2], S. Moutton[1], C. Thauvin-Robinet[1]; [1]Genetics Ctr., FHU-TRANSLAD, Dijon Bourgogne Univ. Hosp., Dijon, France, [2]Inserm UMR1231 team GAD, Univ. of Burgundy and Franche-ComtÃ©, Dijon, France, [3]Neurology Dept., Dijon Bourgogne Univ. Hosp., Dijon, France, [4]Functional Unity of innovative diagnosis for rare diseases, Dijon Bourgogne Univ. Hosp., Dijon, France, [5]Dept. of Internal Med., Competence Ctr. for Autoimmune Cytopenia, Dijon, France, [6]Dept. of Ophthalmology, Eye and Nutrition Res. Group, Dijon Bourgogne Univ. Hosp., Dijon, France, [7]Dept. of Adult Neurophysiology, Dijon Bourgogne Univ. Hosp., Dijon, France, [8]Dept. of Vascular and Interventional Radiology, Image-Guided Therapy Ctr., Dijon Bourgogne Univ. Hosp., Dijon, France

Disclosure Block: **Q. Thomas:** None.

The growing use of next generation sequencing over the last decade has led to a dramatic increase in medical knowledge of Mendelian disorders and also to better awareness of the underestimated phenomenon of multiple diagnoses. The co-occurrence of more than one pathogenic variant in an individual is estimated to occur in up to 5% of patients and may generate different clinical situations depending on the phenotypes associated with these variants and how much they clinically overlap. These situations can be challenging for physicians to recognize, requiring particular attention and knowledge.In this work, we describe the case of a patient referred to our Genetics Department for syndromic gait disorders with childhood onset and a family history of gait disorders. Intertingly, while the proband's phenotype was compatible with complex cerebellar aatxia with upper motor neuron signs and cognitive impairment, her father's phenotype was the one of pure spastic paraplegia. Using exome sequencing, we identified 2 pathogenic missense variants, one in BSCL2 that was inherited from her symptomatic father and one in MT-ATP6 that was inherited from her asymptomatic mother, allowing us to untangle this challenging situation and provide an accurate diagnosis for each family member. While raising awareness on an underappreciated phenomenon, this informative report illustrates the diagnostic challenge that multiple diagnoses can represent and how clinical expertise and exome sequencing can help solve them.

PrgmNr 2184 - An exonic LINE-1 insertion and a novel missense variant in *CC2D2A* identified in siblings with Joubert syndrome using Long-Read Sequencing

View session detail

Author Block: **K. YANAGI**[1], J. Coker[2], K. Miyana[3], M. Omata[1], T. Hidai[1], A. Igarashi[1], K. Satou[1], A. Richman[4], S. Indupuru[4], Y. Matsubara[1], H. Thapar[4], T. Kaname[1]; [1]Natl. Ctr. for Child Hlth.and Dev., Tokyo, Japan, [2]OmniTier, Inc., Rochester, MN, [3]Japanese Red Cross Med. Ctr., Tokyo, Japan, [4]OmniTier, Inc., Milpitas, CA
Disclosure Block: **K. Yanagi:** None.
Joubert syndrome (JBTS) is a genetic and phenotypic heterogeneous disorder characterized by an MRI appearance called 'molar tooth sign', hypotonia in infancy, neonatal breathing dysregulation and developmental delay. *CC2D2A (coiled-coil and C2 domain containing 2A)* is known as a causative gene for JBTS 9 (OMIM 612013). Whole exome analysis by shot-read sequencing could make diagnosis of JBTS. However, there still remain undiagnosed cases. We report that long-read NGS with phasing analysis can reveal the causative biallelic SV and SNV in a family with two Japanese siblings suspected of JBTS. **Case:** The affected siblings, a 24-year-old sister (proband) and a 20-year-old brother, were born to non-consanguineous parents. Both presented hypotonia in infancy, neonatal breathing dysregulation and severe psychomotor developmental delay. They also had an occipital encephalocele. Although JBTS was clinically suspected as one of the different diagnoses, it was not explicitly diagnosed. **Genetic analysis:** We performed whole genome sequencing in the proband using a PacBio Sequel II with CLR (continuous long read) mode. Variant calling displayed an approximately 6.2 kb large insertion in exon 7 and a novel missense variant, c.4454A>G, p.(Tyr1485Cys) in *CC2D2A* (NM_001080522.2). The long insertion sequence was highly similar to LINE-1 (GU477636.1). The affected brother and her mother had the insertion confirmed by long range PCR. The missense variant, which was not recorded in gnomAD and was predicted to be deleterious by SIFT, PolyPhen-2 and other programs, was identified in the brother but not in the mother by Sanger sequencing. Deep sequencing using a MiSeq also could not detect the missense variant in the mother. We could not investigate the variants in the father because he was deceased. Therefore, we performed phasing analysis using SNV calling data from CLR sequences, which clearly showed that the two variants were in *trans* in the patient. SNV calling data around *CC2D2*A from the CLR sequences were verified by target re-sequencing using MiSeq. **Discussion:** We diagnosed and concluded the patients were JBTS affected by the biallelic variants in *CC2D2A*. Our calling program is effective to detect not only SVs but also SNVs even from the CLR sequencing data, which could also be used for phasing analysis. Our analyses can be applicable to identify pathogenic biallelic variants in such patients without samples from parents.

PrgmNr 2186 - Bi-allelic variants in neuronal cell adhesion molecule (*NRCAM*) lead to a novel neurodevelopmental disorder characterized by developmental delay, hypotonia, peripheral neuropathy or spasticity

View session detail

Author Block: **A. Kurolap**[1,2,3], F. Kreuder[4], C. G. Gonzaga-Jauregui[5], M. Plesser Duvdevani[6], T. Harel[6], L. Tammer[7], B. Xin[8], S. Bakhtiari[9,10], J. Rice[11], C. L. van Eyk[12], J. Gecz[13], J. K. Mah[14,15], D. Atkinson[16,17], H. Cope[18], J. A. Sullivan[18], D. Colquhoun[4], J. Henry[19], D. Wlodkowic[19], Y. Parman[20], A. Candayan[21], E. Kocasoy-Orhan[22], A. Ilivitzki[2], S. Soudry[2], R. Leibu[2], F. Glaser[3], V. Sency[23], Undiagnosed Diseases Network, G. Ast[7], V. Shashi[24], M. C. Fahey[4], E. Battaloglu[21], A. Jordanova[25,26], V. Meiner[6], A. M. Innes[27,15], H. Wang[28], O. Elpeleg[6], M. C. Kruer[9,10], J. Kaslin[4], H. Baris Feldman[1]; [1]Tel Aviv Sourasky Med. Ctr., Tel Aviv, Israel, [2]Rambam Hlth.Care Campus, Haifa, Israel, [3]Technion - Israel Inst. of Technology, Haifa, Israel, [4]Monash Univ., Melbourne, Australia, [5]Univ. Natl. AutÃ³noma de MÃ©xico (UNAM), Wautoma, WI, [6]Hadassah Med. Ctr. and Hebrew Univ. of Jerusalem, Jerusalem, Israel, [7]Tel Aviv Univ., Tel Aviv, Israel, [8]DDC Clinic, Ctr. for Special Needs Children, Middlefield, OH, [9]Phoenix Children's Hosp., Phoenix, AZ, [10]Univ. of Arizona Coll. of Med., Phoenix, AZ, [11]Women's and Children's Hosp., Adelaide, Australia, [12]Univ. of Adelaide, Adelaide, SA, Australia, [13]The Univ. of Adelaide, Adelaide, Australia, [14]Alberta Children's Hosp. Res. Inst., Calgary, AB, Canada, [15]Univ. of Calgary, Calgary, AB, Canada, [16]Max Planck Inst. of Immunobiology and Epigenetics, Freiburg, Germany, [17]Univ. of Antwerp, Antwerpen, Belgium, [18]Duke Univ. Med. Ctr., Durham, NC, [19]RMIT Univ., Melbourne, Australia, [20]Istanbul Universtiy, Istanbul, Turkey, [21]Bogazici Univ., Istanbul, Turkey, [22]Istanbul Univ., Istanbul, Turkey, [23]DDC Clinic - Ctr. for Special Needs Children, Middlefield, OH, [24]Duke Univ., Durham, NC, [25]Univ Antwerp - CDE, Antwerpen, Belgium, [26]Med. Univ.-Sofia, Sofia, Bulgaria, [27]Alberta Children's Hosp, Calgary, AB, Canada, [28]DDC Clinic for Special Needs Children, Middlefield, OH

Disclosure Block: **A. Kurolap:** None.

Cell adhesion molecules (CAMs) are membrane-bound proteins predominantly expressed in the central nervous system (CNS) along principal axonal pathways. CAMs play key roles in nervous system development, neural cell differentiation and migration, axonal growth and guidance, myelination, and synapse formation. The most abundant CNS-expressed CAMs are L1CAM, NRCAM, CHL1 and NFASC, all belonging to the L1-CAM subgroup; these proteins are predominantly expressed along principal axonal pathways, such as the corpus callosum, corticospinal tract and the optic nerve. To date, pathogenic variants only in *L1CAM* and *NFASC* have been associated with neurodevelopmental disorders. In this study, we describe 10 patients from eight families with bi-allelic variants in the neuronal cell adhesion molecule *NRCAM*, leading to a novel neurodevelopmental syndrome of varying severity. This syndrome is characterized by developmental delay/intellectual disability, hypotonia, peripheral neuropathy and/or spasticity. Computational analyses of *NRCAM* variants, which mostly cluster in the third fibronectin type III (Fn-III) domain, strongly suggest a deleterious effect on NRCAM protein structure and function, potentially hindering its ability to interact with other proteins. These findings are corroborated by previous *in vitro* studies of murine *Nrcam*-deficient cells, revealing abnormal neurite outgrowth, synaptogenesis and formation of nodes of Ranvier on myelinated axons. We performed studies on zebrafish *nrcama* mutants lacking the Fn-III domains, revealing significantly increased amounts of acetylated Î±-tubulin fibers in the dorsal telencephalon (p=0.02) and the cerebellum (p=0.04), and increased amounts of dp-ERK positive cells in the dorsal telencephalon (p=0.04) and the cerebellum (pnrcama mutant larvae displayed significantly reduced swimming (pNRCAM disruption causes a variable form of a neurodevelopmental disorder, and broadens the knowledge on the role of NRCAM in nervous system development, and the growing role of the cell adhesion molecules family in the nervous system.

PrgmNr 2187 - Comprehensive genetic investigation of penta-nucleotide tandem repeats at *RFC1* locus in Indian ataxia and ALS cohort

View session detail

Author Block: **N. Tyagi**[1], B. Uppili[1], P. Sharma[1], S. Parveen[1], S. Saifi[1], U. Shamim[1], A. Jain[1], S. Sivasubbu[1], V. Scaria[1], A. Nalini[2], A. Srivastava[3], M. Faruq[1]; [1]CSIR-IGIB, New Delhi, India, [2]NIMHANS, Bangalore, India, [3]Dept. of Neurology, AIIMS, New Delhi, India
Disclosure Block: **N. Tyagi:** None.
Monogenic etiologies for cerebellar ataxias and related disorders are ever expanding and till date more than 200 causal genetic loci have been elucidated. A significant numbers of tandem nucleotide repeat expansion associated loci are the prevalent mutations globally. Among these loci, a recently, described biallelic (AAGGG)n expansion in intron 2 of *RFC1* has emerged as a causal loci for CANVAS (Cerebellar ataxia, Neuropathy, and Vestibular areflexia syndrome) and sporadic cases of late onset ataxias. Subsequent studies have linked $(AAGGG)_{exp}$ in association with a common caucasian haplotype. We aimed to determine the occurrence of $(AAGGG)_{exp}$ in a cohort of genetically unknown ataxias and a cohort of Amyotophic Lateral Sclerosis (ALS) and to estimate the *RFC1* disease rik alleles/haplotype in Indian population. A systematic screening of *RFC1*-TNR in ~2000 (ataxia:1908, ALS:476) cases by repeat length dependent PCR selection (flanking PCR, repeat-primed PCR, long-range PCR) of risk alleles and repeat length estimation using short read and long read sequencing. We found three patients with biallelic $(AAGGG)_{exp}$ at *RFC1* i.e. 1/773 late-onset ataxia, 1/95 SCA12-like and 1/476 ALS cases. We found 29 $(AAGGG)_{exp}$ carriers and a new repeat configuration of AAAGGG through next-generation sequencing. We found that the risk haplotype is present in 8% of Indian control population (IndiGen data and 1000 genome SAS population). Study suggests the rarity of biallelic *RFC1*-TNR mutation; observation of a significant % of risk haplotype warrants further study in Indian population among phenotype segregated cohorts e.g. ALS cohort and SCA12 phenocopies.

PrgmNr 2188 - Delineating the spectrum of disorders with CNS white matter abnormalities in Indian population

View session detail

Author Block: **A. Shukla**[1], P. Kaur[1], M. C. do Rosario[1], M. Hebbar[1], S. Sharma[2], N. Kausthubham[1], K. Nair[1], A. Shrikiran[3], R. Bhat Y[3], L. E. S. Lewis[3], S. Nampoothiri[4], S. J. P[5], S. Bijarnia[6], R. D. Puri[7], S. Pai[8], K. Rakshith[8], N. Kamath[9], S. Bajaj[10], A. Kumble[11], R. Shetty[12], R. Shenoy[13], M. Kamate[14], H. Shah[15], M. Muranjan[16], B. L. Yatheesha[17], K. S. Avabratha[18], G. Subramaniam[19], R. Kadavigere[20], S. L. Bielas[21], K. M. Girisha[22]; [1]Dept. of Med. Genetics, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [2]Dept. of Paediatrics, Lady Hardinge Med. Coll. and Associated Kalawati Saran Childrenâ□□s Hosp., New Delhi, India, [3]Dept. of Paediatrics, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [4]Dept. of Paediatric Genetics, Amrita Inst. of Med. Sci. and Res. Ctr., Kochi, India, [5]Div. of Genetics, Mazumdar Shaw Med. Ctr., Narayana Hlth.City, Bangalore, India, [6]Inst. of Med. Genetics and Genomics, Sir Ganga Ram Hosp., New Delhi, India, [7]Sir Ganga Ram Hosp., New Delhi, INDIA, India, [8]Dept. of Neurology, Kasturba Med. Coll., Mangalore, Manipal Academy of Higher Ed., Manipal, India, [9]Kasturba Med. Coll., Manipal Univ., Mangalore, India, [10]Jaslok Hosp. and Res. Ctr., Mumbai, India, [11]Dept. of Paediatrics, Indiana Hosp. and Heart Inst., Mangalore, India, [12]First Neuro Hosp., Mangalore, India, [13]Dept. of Paediatrics, K.S. Hegde Med. Academy, NITTE Univ., Manipal, India, [14]Dept. of Paediatrics, Jawaharlal Nehru Med. Coll., Balgaum, India, [15]Dept. of Orthopaedics, Kasturba Med. Coll., Manipal Academy of Higher Ed., Manipal, India, [16]Dept. of Pediatrics, Genetics Div., Seth Gordhandas Sunderdas Med. Coll. and King Edward VII Mem. Hosp., Parel, Mumbai, India, [17]Dheemahi Child Neurology and Dev. Ctr., Shimoga, India, [18]Dept. of Pediatrics, Father Muller Med. Coll. and Hosp., Mangalore, India, [19]Colours Children Hosp., Nagpur, India, [20]Dept. of Radiodiagnosis, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [21]Univ. of Michigan, Ann Arbor, MI, [22]Dept. of Med. Genetics, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, Manipal, India

Disclosure Block: **A. Shukla:** None.

Introduction: Genetic disorders with predominant central nervous system white matter abnormalities (CNS WMAs), referred to as leukodystrophies, are phenotypically and genetically heterogeneous entities. At least 422 disorders with CNS WMAs have been reported with largely monogenic etiology and a handful of chromosomal and microdeletion duplication syndromes.

Methods: We performed clinical and genetic evaluation of families with CNS WMAs. Targeted genetic testing was performed in families with a clinical diagnosis. Families without a diagnosis after targeted genetic testing or without a clinical diagnosis underwent broad spectrum genetic testing. **Results:** We evaluated 104 individuals with CNS WMAs from 94 unrelated families. Fifty-eight individuals (55.76%) in the present cohort were males and 46 (44.23%) were females. The age ranged from new-born to 59 years, however a majority (n=98, 94.23%) of individuals were of paediatric age. Consanguinity was noted in 55.31% (52/94) of the families. Targeted genetic testing was performed in 13 families with a clinical diagnosis, ten families received diagnosis. Chromosomal microarray (CMA) was performed for two families, one received a diagnosis. Mendeliome sequencing was performed in 10 families, all received diagnosis. Whole exome sequencing (WES) was performed in 77 families and was diagnostic in 51 (66.23%). Singleton exome sequencing was performed in 73/77 families of which 49 (67.12%) families received a genetic diagnosis. Overall, a genetic diagnosis was obtained in 70 families (74.46%). Twenty-two of 42 distinct disorders observed in this cohort have not been reported in Indian individuals previously. Notably, disorders of nuclear mitochondrial pathology were most frequent (8 disorders in 19 families). Thirty-four of 66 (51.51%) disease-causing variants are novel.

Conclusion: The present cohort describes phenotypic and genotypic spectrum of genetic disorders with CNS WMAs in our population. It demonstrates WES, especially singleton WES, as an efficient tool

in diagnosis of these heterogeneous entities. Additionally, it highlights possible founder events and recurrent disease-causing variants in our population and their implications on the testing strategy.

PrgmNr 2189 - Delineation of the phenotypic and genotypic spectrum of disorders with deficient myelination of central nervous system in 26 Indian families

View session detail

Author Block: **M. C. do Rosario**[1], S. Sharma[2], D. L. Narayanan[1], P. H. Somashekar[1], B. L. Yatheesha[3], K. Nair[1], N. Jadhav[1], S. Mishra[1], S. Siddiqui[4], S. J. Patil[5], R. Bhat Y[6], A. Shrikiran[6], R. Kulkarni[7], S. B. Mahay[8], V. Kumar[9], R. Kadavigere[10], P. Upadhyai[1], S. L. Bielas[11], K. M. Girisha[1], A. Shukla[1]; [1]Med. Genetics, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [2]Pediatrics, Lady Hardinge Med. Coll. and Associated Kalawati Saran Childrenâ□□s Hosp., New Delhi, India, [3]Dheemahi Child Neurology and Dev. Ctr., Shimoga, India, [4]Neuroimaging and Interventional Radiology, STAR Inst. of NeuroSci.s, STAR Hosp., Hyderabad, India, [5]Div. of Genetics, Mazumdar Shaw Med. Ctr., Narayana Hlth.City, Bangalore, India, [6]Pediatrics, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [7]Yashwantrao Chavan Mem. Hosp., Postgraduate Inst., Pune, India, [8]Inst. of Med. Genetics and Genomics, Sir Ganga Ram Hosp., New Delhi, India, [9]Pediatric Surgery, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [10]Radiodiagnosis, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [11]Dept. of Human Genetics and Dept. of Pediatrics, Div. of Genetics, Metabolism, and Genomic Med., Univ. of Michigan Med. Sch., Ann Arbor, MI

Disclosure Block: **M.C. do Rosario:** None.

Introduction: Disorders with deficient myelination of central nervous system (CNS) manifest either as permanent hypomyelination, delayed myelination, or often as secondary hypomyelination and are predominantly genetic in origin. Forty-six conditions have been consistently reported with permanent hypomyelination. Of these, 45 are monogenic, presenting predominantly with a recessive mode of inheritance and one attributed to chromosomal abnormalities. Genetic disorders with delayed and secondary hypomyelination are significantly heterogeneous and often associated with impairment of CNS function. **Methods**: We performed clinical and genetic evaluation of Indian families with deficient myelination of CNS. Families with clinical diagnoses of disorders without genetic heterogeneity and known to be caused by recurrent variants/variants in small genes underwent targeted genetic testing. Families with genetically heterogenous disorders, defects in large genes, no clinical diagnosis, or undiagnosed by targeted genetic testing underwent broad spectrum genomic testing. **Results**: We evaluated 31 individuals with deficient CNS myelination from 26 unrelated families. Twenty individuals (64.5%) are males and 11 (35.5%) females with age range of four months to 15 years. Consanguinity was noted in 53.8% (14/26) families. Of the six families with clinical diagnoses, 4 (66.66%) received genetic diagnosis by targeted testing. Twenty-two families including two families undiagnosed by targeted genetic testing, underwent genomic testing. Of these, 15 (71.42%) families received diagnosis by exome sequencing, and 1 (14.28%) by chromosomal microarray. Overall, 20 families had genetic diagnoses (76.92%). Ten families (50%) had disorders associated with permanent hypomyelination, 2 (10%) had secondary hypomyelination and 8 (40%) had delayed myelination. We observed 17 disorders in 20 families, of which 5 (29.4%) are previously not reported in Indian population. We observed 21 variants, of which 15 are single nucleotide variants and four are copy number variants. Of these, 52.38% (11/21) are homozygous. Ten (47.61%) variants in this cohort are novel. **Conclusion**: This ongoing study aims to define clinical and genomic spectrum of genetic disorders with deficient CNS myelination in India and elucidate pathomechanisms underlying these disorders. The study lays groundwork for diagnosis and possible therapeutic interventions for Indian families with these disorders.

PrgmNr 2190 - Expanding the genotypic spectrum of Allan-Herndon-Dudley syndrome: Report of two novel mutations in the *SLC16A2* gene

View session detail

Author Block: **M. Santos Melo**[1], S. Lemos Ferreira[1], J. Catanho da Silva[1], S. Santos[2], M. Paiva[2], C. Limbert[3], S. Jacinto[4], A. Ferreira[5], D. Antunes[1]; [1]ServiÃ§o de GenÃ©tica MÃ©dica, Hosp. Dona EstefÃ¢nia, Centro Hosp.ar e UniversitÃirio de Lisboa Central, Lisboa, Portugal, [2]ServiÃ§o de Pediatria MÃ©dica, Hosp. Dona EstefÃ¢nia, Centro Hosp.ar e UniversitÃirio de Lisboa Central, Lisboa, Portugal, [3]Unidade de Endocrinologia, ServiÃ§o de Pediatria MÃ©dica, Hosp. Dona EstefÃ¢nia, Centro Hosp.ar e UniversitÃirio de Lisboa Central, Lisboa, Portugal, [4]ServiÃ§o de Neuropediatria, Hosp. Dona EstefÃ¢nia, Centro Hosp.ar e UniversitÃirio de Lisboa Central, Lisboa, Portugal, [5]Centro de ReferÃªncia de DoenÃ§as HereditÃirias do Metabolismo, Hosp. Dona EstefÃ¢nia, Centro Hosp.ar e UniversitÃirio de Lisboa Central, Lisboa, Portugal

Disclosure Block: **M. Santos Melo:** None.

Introduction: Allan-Herndon-Dudley syndrome (AHDS, OMIM #300523) is an X-linked disorder caused by hemizygous pathogenic variants in the *SLC16A2* gene. Clinically, AHDS is characterized by neurologic findings, dysthyroidism and pathognomonic thyroid test results (high free T3 and low reverse T3, low free T4, and normal or slightly elevated TSH). The neurocognitive phenotype includes hypotonia, feeding difficulties in infancy, developmental delay, intellectual disability ranging from mild to severe, pyramidal signs, extrapyramidal findings, and late-onset seizures. Dysthyroidism signs include poor weight gain, reduced muscle mass, cold intolerance, sweating, elevated heart rate and irritability. To date, over 100 different genetically confirmed cases were reported. Here, we present two newly identified patients harboring two novel *SLC16A2* variants, aiming to expand the AHDS clinical and mutational spectrum. **Case reports**: Two unrelated boys were referred for genetic evaluation at 3 and 10 years of age, respectively. Both patients presented with severe global developmental delay, hypotonia, absence speech, and dystonia. The older child also had pyramidal signs, and dysthyroidism features including feeding difficulties, poor weight gain, reduced muscle mass and variable cold intolerance. Family history was suggestive of an X-linked inheritance pattern. Clinical exome sequencing revealed two hemizygous variants: *SLC16A2* (NM_006517.3) c.1057A>C p.(Thr353Pro) in proband 1, and *SLC16A2* (NM_006517.3) c.695C>T p.(Ser232Phe) in proband 2, classified as variants of unknown significance. Both variants were neither present in population databases nor reported in the literature. *In silico* analysis suggested that variants were disease causing. Segregation analysis revealed that mother and grandmother were carriers in both families. Further, retrospective evaluation confirmed abnormal thyroid function: normal TSH, low free-T4, high free T3 and low reverse T3. With the available evidence, we reclassified both variants as likely pathogenic, establishing the diagnosis of AHDS. **Conclusions**: Here, we identified two previously unreported missense variants in *SLC16A2* gene. Both patientsâ phenotypes were consistent with other reported cases. Reverse phenotyping through thyroid hormonal profile and segregation analysis were essential for confirmation of variants pathogenicity and establishment of molecular diagnosis. Additionally, this report further broadens the AHDS genotypic spectrum, and provides additional information for genotype-phenotype correlations.

PrgmNr 2191 - Genotypic spectrum and its clinical implication in disorders with epilepsy in Indian population: A preliminary experience

View session detail

Author Block: **P. Majethia**[1], P. H. Somashekar[1], K. Nair[1], N. Jadhav[1], D. L. Narayanan[1], S. Mishra[1], A. Shrikiran[2], R. Bhat Y[2], S. Sharma[3], B. K. Praveen[4], B. L. Yatheesha[5], S. J. Patil[6], N. Kamath[7], S. Siddiqui[8], R. Kadavigere[9], S. L. Bielas[10], K. M. Girisha[1], A. Shukla[1]; [1]Med. Genetics, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [2]Pediatrics, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [3]Neurology Div., Dept. of Pediatrics, Lady Hardinge Med. Coll. and Associated Kalawati Saran Children's Hosp., New Delhi, India, [4]Pediatrics, Father Muller Med. Coll. Hosp., Mangalore, India, [5]Paediatric neurology, Dheemahi Child Neurology and Dev. center, Shimogga, India, [6]Pediatrics, Narayana Hrudayalaya Hosp./ Mazumdar-Shaw Med. Ctr., Bangalore, India, [7]Pediatrics, Kasturba Med. Coll., Mangalore, Manipal Academy of Higher Ed., Manipal, India, [8]Neuroimaging and Interventional Radiology, STAR Inst. of NeuroSci.s, STAR Hosp., Hyderabad, India, [9]Radio Diagnosis, Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [10]Dept. of Human Genetics and Dept. of Pediatrics, Div. of Genetics, Metabolism, and Genomic Med., Univ. of Michigan Med. Sch., Ann Arbor, MI

Disclosure Block: **P. Majethia:** None.

Introduction: Disorders with epilepsy are genetically and phenotypically heterogeneous. The rapid advancement of genetic testing has helped address these complexities and aided definitive diagnosis, precision medicine and genetic counseling in families with these disorders. **Methods:** We recruited 49 families (51 individuals) with neurodevelopmental disorder (NDDs) with epilepsy. Four individuals (four families) with clinically recognizable phenotypes underwent targeted genetic testing and 48 individuals (46 families) with clinically unrecognizable phenotypes and/or undiagnosed by targeted testing underwent genomic testing after detailed clinical evaluation. The implications on genetic counseling and therapy were evaluated in individuals with definitive diagnosis. **Results:** Our cohort comprises of 24 males (47%) and 27 females (53%) from 19 consanguineous (39%) and 30 non-consanguineous (61%) families. A definitive molecular diagnosis was achieved in 32 families with three being diagnosed by targeted testing and 29 by exome sequencing (ES). Twenty-nine monogenic disorders and one imprinting/microdeletion syndrome (Angelman syndrome) were identified. Of these, eight (25%) families had variants in seven genes causing developmental epileptic encephalopathies (*KCNQ2*, *STXBP1 UGDH*, *FGF13*, *AP3B2*, *FGF12*, *CYFIP2*). Seven (23%) families had variants in seven genes causing metabolic disorders (*HEXB, GCDH, GCSH, PNPT1, SHMT2, CARS2, SLC25A10*) of which six had variants in six nuclear encoded mitochondrial genes. Seven (23%) families had variants in 6 genes causing disorders with epilepsy as a core symptom (*SCN1A, KCTD7, NAXD, TRAPPC4, TPP1, KCNJ10*) and nine (29%) had variants in nine genes causing NDDs associated with epilepsy (*KMT2A, AP4S1, DYRK1A, RNASEH2C, DYNC1H1, GALNT2, ARID1B, ANKRD11, MECP2*). Biallelic variants were identified in 17 families (53%), *de novo* in 13 (41%), and heterozygous exonic deletion in one (3%) family. Seventeen (55%) of 31 disease-causing variants were novel. We also report variants in *SHMT2, SLC25A10*, and *FGF13* causing extremely rare disorders with less than 10 patients reported worldwide. Prenatal diagnosis was carried out in 19% (5/26) of families. Notably, ES had therapeutic implications in 50% of individuals (13/26) with a definitive diagnosis. **Conclusion:** The above-mentioned cohort is a part of an ongoing study of disorders with epilepsy. We herein describe the first cohort elucidating the genotypic spectrum and its clinical implications in disorders with epilepsies in Indian population.

PrgmNr 2192 - Identification of *NDEL1* as a novel gene for lissencephaly

View session detail

Author Block: **M-H. TSAI**[1,2], H-C. Ke[3], C-Y. Fong[4], C-C. Ng[4], K-S. Lim[4], J-W. Tsai[3]; [1]Kaohsiung Chang Gung Mem. Hosp., Kaohsiung, Taiwan, [2]Chang Gung Univ., Taoyuan, Taiwan, [3]Natl. Yang-Ming Ciao-Tung Univeristy, Taipei, Taiwan, [4]Univ. of Malaya, Kuala Lumpur, Malaysia
Disclosure Block: **M. Tsai:** None.
Background Lissencephaly (LIS) is a rare neurological disorder caused by abnormal neuronal migration. More than 20 LIS genes have been reported so far, most of them are functionally associated with the cytoskeleton. In a recent large study, more than 80% of patients with lissencephaly can have a pathogenic variant in reported genes. The genetics of lissencephaly in Asian patients is relatively under-explored. In addition, novel genes remain to be identified in unsolved cases. Method We recruited 33 lissencephaly patients and their family members. Candidate gene Sanger sequencing or whole exome sequencing (WES) was performed to search for known genetic causes. Trio WES study was used to study unsolved cases. In utero electrophoresis model was performed to investigate the effect of NDEL1 mutant on neuronal migration. Results In total, 33 patients with lissencephaly were enrolled in this study. 20/33 (60.6%) had a pathogenic variant in reported genes, including 7 DCX, 2 CEP85L, 2 DYNC1H1, 2 MAST1, 2 chromosome 17 deletion, 1 PAFAH1B1, 1 WDR62, 1 TUBA1A, 1 GRIN1, and 1 BICD2. WES Trio study was performed on the remaining 13/33 (39.3%) patients. One patient with a de novo NDEL1 missense variant was identified. NDEL1 is known to interact with LIS1 (PAFAH1B1) and required for microtubule organization and anchoring microtubule to the centrosome. Our preliminary functional study demonstrated that NDEL1 knock down and mutant both impaired neuronal migration. Conclusion The currently known lissencephaly genes are accounted for ~60% of our Asian cohort. We reported that NDEL1 is a novel gene for lissencephaly in human. NDEL1 dysfunction impaired neuronal migration during brain development.

PrgmNr 2193 - LRFN1- a novel candidate gene as a potential cause of autism spectrum disorder

View session detail

Author Block: **J. Shah**[1], D. Jain[2], F. J. Sheth[1], H. Sheth[1]; [1]FRIGE's Inst. of Human Genetics, Ahmedabad, India, [2]Shishu Child Dev. and Early Intervention Ctr., Ahmedabad, India
Disclosure Block: **J. Shah:** None.
Autism spectrum disorder (ASD), a complex neurodevelopmental condition is broadly characterized by impaired social interactions and stereotypy. It has an estimated prevalence of â¥64 per 10,000 people worldwide. Genetic factors have been recently identified which are associated with ASD phenotype, although the genetic etiology is highly variable and explains only 30-40% of the cases. Whole exome sequencing has led to the identification of several novel genes for ASD. Considering 10-15% cases with a confirmed genetic diagnosis have a *de novo* mutation, many novel genes have been identified using a patient-parentsâ trio based approach. We report the utility of this approach in a child that was clinically diagnosed with ASD according to DSM-V criteria. WES in the proband showed a heterozygous, missense variant c.176T>C (p.V59A) in exon 1 of the *LRFN1* gene (ENST00000248668.4). Validation and segregation analysis by Sanger sequencing showed the variant to be *de novo* in origin and classed as likely pathogenic according to the ACMG-AMP classification system. *LRFN1* is predominantly expressed in adult and fetal brain and is involved in the promotion of neurite outgrowth in hippocampal neurons, regulation and maintenance of excitatory synapses and induction of clustering of excitatory postsynaptic proteins, including DLG4, DLGAP1, GRIA1 and GRIN1. Literature has suggested a strong association between synaptic proteins and autism. Furthermore, mutations in genes encoding the aforementioned postsynaptic proteins are known to be associated with ASD. Collectively, the *in-silico* information, gene expression and interaction pathway along with mouse expression studies suggests *LRFN1* gene as a potential novel candidate gene associated with ASD.

PrgmNr 2194 - Metabotropic glutamate receptor subtype 2 (*GRM2*) is a novel disease gene for developmental and epileptic encephalopathy

View session detail

Author Block: **C. W. Lam**; The Univ. of Hong Kong, Hong Kong, China
Disclosure Block: **C.W. Lam:** None.
Introduction: Developmental and epileptic encephalopathy (DEE) is not an uncommon condition and currently 96 genes causing DEE have been reported in the literature. Most of the DEE genes are neuronal channels or receptors. With the advance of whole exome sequencing (WES), the underlying genetic defects in patients with DEE can be demystified. Here, we identified a *de novo* novel missense variant in the *GRM2* gene (metabotropic glutamate receptor subtype 2) in a patient with DEE. **Case:** We report a case of infantile-onset epilepsy in a 6 months old Chinese girl. The patient was first presented at birth with twitching movements and was later referred to us for convulsion of unknown aetiology. EEG showed hypsarrhythmia, generalized slow spike and wave discharges. Biochemical investigations were normal and there was no family history of similar disorders. Developmental delay became apparent only after the onset of seizures. The patient was diagnosed to have DEE. **Method:** Clinical genetic and genomic analyses included conventional PCR and Sanger sequencing and WES. The bioinformatics analysis was performed using in-house pipelines. **Results:** No known pathogenic variant was found in the known disease-causing genes for DEE. A novel heterozygous missense variant, NM_000839: c.2582G>A, p.Gly861Asp was found on the last exon of the *GRM2* gene. The c.2582G>A variant is an unreported variant and not found in population databases (ExAC and GnomAD). The variant was undetected in the parents. According to the American College Medical Genetics Guideline 2015 for the interpretation of sequence variants, the variant c.2582G>A was classified to be likely pathogenic. **Conclusions:** L-glutamate is the major excitatory neurotransmitter in the brain and activates metabotropic glutamate receptors. Though GRM2 is important in glutamatergic neurotransmission, the role of GRM2 in epilepsy and development has not been studied. We envisage metabotropic glutamate receptors may be potential therapeutic targets for epilepsy control.

PrgmNr 2195 - Mono- and bi-allelic variation of *ABCA2* in individuals with different neurodevelopmental disorders

View session detail

Author Block: **K. Ounap**[1,2], K. Reinson[1,2], K. T. Oja[2], E. G. Seaby[3], M. H. Wojcik[3,4], A. Oâ□□Donnell-Luria[3,4], B. M. Wittmann[5], A. D. Kennedy[5], C. Zweier[6,7], G. Vasileiou[6], G. N. Wilson[8], I. A. Alkhawaja[9], R. L. Sanchez Russo[10], T. Wirth[11], S. Syrbe[12], K. Platzer[13], S. Alyamani[14], F. Alkuraya[14], S. Pajusalu[1,2]; [1]Dept. of Clinical Genetics, United Lab., Tartu Univ. Hosp., Tartu, Estonia, [2]Inst. of Clinical Med., Univ. of Tartu, Tartu, Estonia, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Boston Childrenâ□□s Hosp., Boston, MA, [5]Metabolon, 615 Davis Drive, Suite 100, Morrisville, NC, [6]Inst. of Human Genetics, Friedrich-Alexander-Univ. Erlangen-NÃ¼rnberg, Erlangen, Germany, [7]Dept. of Human Genetics, Univ. Hosp. Inselspital, Univ. of Bern, Bern, Switzerland, [8]Texas Tech Univ. Hlth.Sci. Ctr., Lubbock and KinderGenome Med. Genetics, Dallas, TX, [9]Albashir Hosp., Amman, Jordan, [10]Dept. of Human Genetics, Emory Univ., Atlanta, GA, [11]Strasbourg Univ. Hosp., Strasbourg, France, [12]Div. of Pediatric Epileptology, Ctr. for Paediatrics and Adolescent Med., Univ. Hosp. Heidelberg, Heidelberg, Germany, [13]Inst. of Human Genetics, Univ. of Leipzig Med. Ctr., Leipzig, Germany, [14]King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia

Disclosure Block: **K. Ounap:** None.

The ATP binding cassette transporter ABCA2, encoded by the *ABCA2* gene, is an endolysosomal membrane protein and has a critical role for mediating the movement of sphingolipids within cellular compartments and maintaining homeostasis of sterols, sphingolipids and cholesterol. ABCA2 is most highly expressed in brain tissue and knockout mice display neurological defects consistent with aberrant myelination. We collected data from nine patients in eight unrelated families and added four cases from the literature. Five patients had *de novo* heterozygous variants (HTZ) and eight bi-allellic variants [seven homozygous variants (HMZ) and one compound heterozygous variant (cHTZ)] in *ABCA2*. All but one of the patients (with the cHTZ) had delayed psychomotor development. Two patients (one HMZ; one cHTZ) did not develop intellectual disabilities (ID); the remaining 11 had mild to moderate ID. Brain MRI was performed on eight patients - four had normal (two HTZ; two HMZ) results and the remainder had different abnormalities, but none had aberrant myelination as observed in the *ABCA2* knockout mice. Autism spectrum disorder occurred in three HTZ and one HMZ patient. Epilepsy was diagnosed in one patient with HTZ and five patients with HMZ. Different types of movement disorders (e.g. dystonia, ataxia) were diagnosed in five HMZ, two HTZ and one cHTZ. No patients were reported to have abnormal levels of serum cholesterol and/or lipoproteins, except two HMZ brothers with elevated level of low-density lipoprotein. However, untargeted metabolomic analysis performed from plasma revealed low levels of many sphingolipids in a patient with a *de novo* missense variant. The phenotype of the collected patients was variable with the core features of ID and various types of movement disorders in most patients. No clear correlation between the phenotype and the different inheritance patterns was observed. Various variants in the *ABCA2* gene were reported, such as missense variants, deletions or duplications; and they were located in different parts of the gene, with no clear correlation between the location or the type of the variant and clinical features. These findings suggest *ABCA2* to be associated with a spectrum of neurodevelopmental disorders, but highlights the need to further collect and investigate patients with variants in *ABCA2*.

Funding: Estonian Research Council grants PUT355, PRG471, MOBTP175; Centers for Mendelian Genomics grant NHGRI U01HG008900.

PrgmNr 2196 - Multiple methodologies reveal novel picture of the genetic architecture of intellectual disability in Northern Finland

View session detail

Author Block: **L. Urpa**[1], M. Kurki[1,2,3], E. Saarentaus[1], O. PietilÃ¤inen[2,4,5], E. HÃ¤mÃ¤lÃ¤inen[1], V. Salomaa[6], J. Suvisaari[6], E. Rahikkala[7,8,9,10], R. Keski-Filppula[7,8,9], M. Rauhala[11], S. Korpi-HeikkilÃ¤[11], J. Komulainen-Ebrahim[7,12], H. Helander[7,12], P. Vieira[7,12], J. Moilanen[7,8,9], J. KÃ¶rkkÃ¶[11], O. Kuismin[1,7,8,9], M. J. Daly[1,2,3,13,14], A. Palotie[1,2,3,13,14]; [1]Inst. for Molecular Med. Finland, Univ. of Helsinki, Helsinki, Finland, [2]The Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [4]Dept. of Stem Cell and Regenerative Biology, Harvard Univ., Cambridge, MA, [5]HiLIFE NeuroSci. Ctr., Univ. of Helsinki, Helsinki, Finland, [6]Finnish Inst. for Hlth.and Welfare, Helsinki, Finland, [7]PEDEGO Res. Unit, Univ. of Oulu, Oulu, Finland, [8]Med. Res. Ctr., Oulu Univ. Hosp., Univ. of Oulu, Oulu, Finland, [9]Dept. of Clinical Genetics, Oulu Univ. Hosp., Oulu, Finland, [10]Inst. of Biomedicine, Univ. of Turku, Turku, Finland, [11]Northern Ostrobothnia Hosp. District, Ctr. for Intellectual Disability Care, Oulu, Finland, [12]Dept. of Children and Adolescents, Oulu Univ. Hosp., Med. Res. Ctr. Oulu, Univ. of Oulu, Oulu, Finland, [13]Analytic and Translational Genetics Unit, Dept. of Med., Massachusetts Gen. Hosp., Boston, MA, [14]Dept. of Neurology, Massachusetts Gen. Hosp., Boston, MA

Disclosure Block: **L. Urpa:** None.

Intellectual disability (ID) is a highly heterogenous disorder, yet analysis of the genetic etiology of ID often focuses on only a single mechanism (e.g. *de novo* or pedigree analysis) and excludes others. In the Northern Finland Intellectual Disability (NFID) study, we analyzed de novo variants in trios, inherited variants in large pedigrees, and the polygenic contribution to ID in 901 families (352 trios). We hypothesized that families with a single case of ID (enriched in our cohort for mild ID) were more likely to be explained by *de novo* variants, and families with multiple cases were more likely to be explained by inherited variants and polygenic burden. In *de novo* analysis, we compared the number of observed loss-of-function (LOF) mutations in known ID-associated genes to that expected by known mutation rates, which is well-powered with small sample sizes. We found that there were significantly more LOF mutations in known ID genes than expected, but only in families with a single case of ID (1.6 expected, 30 observed, p=1.34e-20, n=243, 46.7% mild ID) and not in families with multiple cases of ID (0.5 exp., 1 obs., p=0.398, n=109, 34.7% mild ID). When we examined the proportion of these families with an inherited rare, damaging variant in a known ID-associated gene, we found it was similar between families with a single ID case (43/243, 17.7%) and families with multiple ID cases (15/109, 12.8%, p=0.277). This suggests alternative modes of inheritance or damaging variants in novel genes in these multiply-affected families. When we investigated the polygenic contribution to ID, we saw that all ID cases had a lower polygenic score (PGS) for cognitive performance (CP; mean=-0.098 [95%CI:-0.165, -0.031], p=4.77e-3) and educational attainment (EDU; mean=-0.111 [95%CI:-0.178, -0.044], p=1.54e-3, n=901) compared to population controls, which persisted in ID cases from families with a single affected individual (CP: mean=-0.102 [95%CI:-0.172,-0.032] p=4.71e-3; EDU: mean=-0.118 [95%CI:-0.188,-0.047], p=1.35e-3, n=815, 47.7% mild ID). In ID cases from families with multiple affected individuals we did not see a difference in PGS from population controls (CP: mean=-0.058 [95%CI:-0.270,0.153], p=0.586; EDU: mean=-0.053 [95%CI:-0.265,0.159], p=0.641; n=86, 53.4% mild ID), but considering the small n we interpret this finding with caution. Over all, we begin to paint a picture of genetic architecture in our cohort that hints at possible different genetic etiologies in spontaneously-occurring vs inherited ID. Future work will focus on finding novel ID-associated genes and missing etiology in our large pedigrees.

PrgmNr 2197 - Mutational Landscape of a Bangladeshi Cohort of Neurodevelopmental Disorders

View session detail

Author Block: **H. Akter**[1,2], S. Shaoli[3,4], M. Basiruzzaman[5,6], I. Mazharul[3,6], R. Md.[1], S. Hossain[2], T. Eshaque[1], R. Md. Ashiquir[5], N. Dity[1], A. Md.[7], H. Md Monir[6], M. Lopa[5], N. Jahan[5], A. Islam[1,8], S. Hasan[8], S. Kanta[4], R. Kazal[9], R. Sumi[10], A. Ahmed[11], Z. Ashraf[11], N. Nassir[11], K. Fatema[12], A. Nabi[2], K. Uddin[1,13], M. Uddin[11,8]; [1]Genetics and Genomic Med. Ctr., NeuroGen Childrenâ□□s Hlth.care, Dhaka, Bangladesh, [2]Dept. of Biochemistry and Molecular Biology, Univ. of Dhaka, Dhaka, Bangladesh, [3]Dept. of Child Neurology, NeuroGen Childrenâ□□s Hlth.care, Dhaka, Bangladesh, [4]Dept. of paediatric neuroSci., dhaka Shishu Hosp., Dhaka, Bangladesh, [5]Ctr. for Precision Therapeutics, NeuroGen Childrenâ□²s Hlth.care, Dhaka, Bangladesh, [6]Dept. of Paediatric Neurology, Natl. Inst. of NeuroSci. and Hosp., Dhaka, Bangladesh, [7]Dept. of Med., Dhaka Med. Coll., Dhaka, Bangladesh, [8]Cellular Intelligence Lab, GenomeArc Inc.,, Toronto, ON, Canada, [9]Dept. of Obstetrics & Gynaecology, Bangabandhu Sheikh Mujib Med. Univ., Dhaka, Bangladesh, [10]Sheikh Hasina Natl. Inst. of Burn & Plastic Surgery, Dhaka, Bangladesh, [11]Coll. of Med., Mohammed Bin Rashid Univ. of Med. and Hlth.Sci., Dubai, United Arab Emirates, [12]Dept. of Paediatric Neurology, Bangabandhu Sheikh Mujib Med. Univ., Dhaka, Bangladesh, [13]Dept. of Biochemistry, Holy Family Red Crescent Med. Coll., Dhaka, Bangladesh

Disclosure Block: **H. Akter:** None.

Background: Single nucleotide variations (SNVs) and copy number variations (CNVs) play a critical role into the pathogenesis of neurodevelopmental disorders (NDDs). In this study, we have applied whole exome sequencing (WES) and chromosomal microarray (CMA) to detect clinically relevant variations/genes in a Bangladeshi NDD cohort.

Methods: We have conducted WES (33.05 Mb human coding regions) and genome-wide CMA (642,824 probes spanning the genome) analysis for 375 NDD patients to identify SNVs and CNVs.

Results: The cohort comprises 68.53% male and 31.47% female, age ranges from 9 days to 31 years where male to female ratio is 2.2:1. Of the total cohort (375 cases), CMA, WES and combined test (CMA and WES) was done on 56.53% (212), 32.00% (120) and 11.47% (43) cases, respectively. The diagnostic yield from CMA, WES and combined test for broader NDD is 12.26% (26/212), 20.00% (24/120) and 32.56% (14/43), respectively. Moreover, we have found 47.17% (100/212), 50.00% (60/120) and 48.84% (21/43) patients carrying variant of uncertain significance (VOUS) in CMA, WES and combined test respectively. An increased diagnostic yield was observed from WES 37.5% (9/24) and combined test 70.00% (7/10) in patients with epilepsy. All the epilepsy patients carrying a pathogenic mutation from our WES analysis are CNV negative in the combined test cohort confirms the utility of WES test for epilepsy cases. In the CMA cohort, we have found a significantly greater head circumference in male compared to female (p=0.0002). Analysis of both pathogenic and VOUS CNVs for â□□critical exonâ□□ genes (CEGs) yielded 153 unique CEGs in pathogenic CNVs and 31 in VUS. Of the unique 184 CEGs, we have considered *PSMC3* gene as a potential candidate gene for the NDD. 95 NDD patients were evaluated for autism spectrum disorders by Autism Diagnostic Observation Schedule-Second Edition (ADOS-2) method. Analyzing the ADOS-2 score, a trend of increased impairment in social communication (p=0.041) was observed in patients carrying duplication CNV compared to patients carrying deletion CNV. Rest of the data analysis of the WES and combined cohort are ongoing.

Conclusions: Our findings demonstrate the utility of WES and CMA as a first-tier diagnostic approach in cases lacking a clear differential diagnosis in a developing country with limited healthcare resources. The results also highlight the utility of WES for precise genetic diagnosis of epilepsy and its integration into the diagnosis therapeutics and management of epilepsy patients. To our knowledge, this is the first genomic research using larger number of NDD patients of Bangladesh.

PrgmNr 2198 - Novel *PCDH19* frameshift variant in a girl with recurrent seizures and phenotype overlapping with *SCN1A* gene dysfunction: a case report and literature review

View session detail

Author Block: **M. Mijovic**[1], J. Ruml Stojanovic[1], A. Miletic[1], B. Bosankic[1], H. Petrovic[1], G. Cuturilo[1,2]; [1]Univ. Children's Hosp., Dept. of Clinical Genetics, Belgrade, Serbia, [2]Faculty of Med., Univ. of Belgrade, Belgrade, Serbia
Disclosure Block: **M. Mijovic:** None.
Case report: We report the case of a 19-month-old girl with recurrent seizures from the nine months of age. Most seizures were provoked by fever and acute illness. Her EEG showed generalized discharges with specific graph elements. Brain MRI was normal. Early psychomotor development in the first year of life was orderly. From the second-year parents noticed a developmental delay, predominantly poor expressive speech development and attention deficit. Prenatal and perinatal anamnesis were unremarkable. Family history regarding seizures and intellectual disability is negative. Clinically, there was a strong suspicion of a disorder from a phenotypic spectrum of *SCN1A* gene dysfunction. Materials and methods: We performed whole exome sequencing and identified causative novel heterozygous *PCDH19* gene variant. In addition, we done a literature research regarding pediatric female cases of *PCDH19* gene dysfunction. Aim of present study was to establish the similarity between phenotypes of *PCDH19* and *SCN1A* gene disfunction. Results: Heterozygous variant identified in the *PCDH19* gene (c.243_246dup, p.Thr83Glyfs7*) creates a shift in the reading frame starting at codon 83, with new reading frame ending in a stop codon 6 positions downstream. It is classified as a likely pathogenic variant according to the ACMG recommendations. Pathogenic variants in the in the *PCDH19* gene, have been associated with developmental and epileptic encephalopathy type 9 (OMIM: 300088), an intriguing X-linked disorder affecting heterozygous females and sparing hemizygous males. Conclusion: Literature data suggests that additional clinical signs, like behavioral difficulties such as autistic spectrum disorders may rather correlate with *PCDH19* than with *SCN1A* gene dysfunction. Nevertheless, there is almost complete overlap in neurodevelopmental and epileptic phenotype in females with *PCDH19* and *SCN1A* gene alteration presented in infancy and early childhood. Distinction between these diseases is important for the proper therapy as well as recurrence risk estimation, since *PCDH19* gene dysfunction can be transmitted through asymptomatic fathers. This suggests the importance of multigene testing based on NGS sequencing in children, especially female, with a clinical diagnosis of Dravet or Dravet-like syndrome.

PrgmNr 2199 - Preliminary studies on apparent Mendelian psychotic disorders in consanguineous families

View session detail

Author Block: **A. Kanwal**[1], S. Sheikh[2], A. Iftikhar[3], S. Naz[4], J. V Pardo[5,6]; [1]Sch. of Biological Sci., Univ. of the Punjab, Lahore, Pakistan, [2]Psychiatry Dept. Hawkes Bay DHB, Hastings, New Zealand, [3]Rainbow obesity and eating Disorders Ctr., Lahore, Pakistan, [4]Univ. of the Punjab, Lahore, Pakistan, [5]Dept. of Psychiatry, Univ. of Minnesota, Minneapolis, Minnesota, USA, MN, [6]Minneapolis VA Hlth.Care System, Minneapolis, Minnesota, USA, MN

Disclosure Block: **A. Kanwal:** None.

A psychiatric disorder is a mental condition in which emotions, cognition, behavioral patterns and moods of an individual are disturbed. Multiple psychiatric disorders recur in families, suggesting a role of genetic factors. Genome-wide association studies have identified these complex traits as highly polygenic. We hypothesize that variants, even though very rare, may exist associated with psychiatric disorders following Mendelian inheritance. Some of these causative gene variants can be identified in families with multiple affected individuals born to unaffected consanguineous parents. Here, we focused on the phenotype of psychosis, since the symptom is severe, demands clinical attention, and can occur in several Diagnostic and Statistical Manual of Mental Disorders (DSM) such as schizophrenia, bipolar disorder, substance abuse, and dementia. We visited outpatient departments and psychiatric wards of multiple hospitals in Lahore, Pakistan. We contacted families with individuals apparently inheriting psychosis in an autosomal recessive or X-linked recessive mode. Patients were first diagnosed by psychiatrists at the hospitals. Later, we administered detailed clinical assessments using Diagnostic Interview for Genetic Studies (DIGS), Diagnostic Interview for Psychosis and Affective Disorders (DI-PAD), Positive and Negative Syndrome Scale (PANSS), Hamilton Depression Rating Scale (HAM-D) and Hamilton Anxiety Rating Scale (HAM-A) standardized rating tools to all affected and unaffected participants. These rating tools confirmed rigorously the diagnosis of the patients, as well as the severity of their symptoms. Similarly, these assessments corroborated the unaffected status of the participating relatives of the patients. In total, we identified five families with ten individuals suffering from schizophrenia, three families with six individuals diagnosed with bipolar disorder and one family with two patients exhibiting unspecified psychosis. All affected individuals were born to unaffected consanguineous parents, most of whom were first cousins. We have obtained DNA samples from consenting affected and unaffected individuals of the families for future genetic analyses. We believe our continued research will help in the identification of rare recessively inherited gene variants for psychiatric disorders. These findings could pinpoint underlying biological mechanisms altered in these devastating diseases. This information will be useful in developing targeted therapies in future. Our work is funded by NIMH, NIH USA, 1R21MH120692-01A1 (SN & JVP).

PrgmNr 2200 - Similar genotypes leading to different phenotype: expansion of the phenotype of UDP-Glucose-6-Dehydrogenase recessive neurodevelopmental disorder with or without epileptic encephalopathy, 2 new descriptions

View session detail

Author Block: **P. Plante**[1], S. Boussion[1], M. Rama[1], P. Brunelle[1], C. Thuillier[1], C. Vanlerberghe[2,3], R. Caumes[2,3], C. Colson[2,3], J. Ghoumid[2,3], T. Smol[1,3]; [1]Inst. de GÃ©nÃ©tique MÃ©dicale, CHU Lille, Lille, France, [2]Clinique de GÃ©nÃ©tique, CHU Lille, Lille, France, [3]EA7364 RADEME, UniversitÃ© de Lille, Lille, France
Disclosure Block: **P. Plante:** None.
Combination of loss-of-function *UGDH* alleles have been recently described in autosomal-recessive cases of early infantile epileptic encephalopathies (EIEE). Most of the reported patients presented pharmaco-resistant epilepsy, severe intellectual disability (ID) with absent speech, significant motor delay, and structural abnormalities on cerebral MRI. Here, we report on two new cases of *UGDH* syndrome expanding the phenotype, with one case without any seizure history.
P1 is a 2 y/o boy with known *UGDH* syndrome phenotype, ie pharmacoresistant EIEE, axial hypotonia, pyramidal signs, microcephaly, no development of motor milestones, absent speech, feeding difficulties, abnormal movements without structural abnormalities on cerebral MRI. P2 is an 8 y/o girl presenting with severe global developmental delay (walk at 3), no speech, stereotypic movements, sleep disturbances and scoliosis. She has no feeding difficulties and no seizures history.
Both patients benefited from exome trio sequencing considering their developmental delay. Compound heterozygous variants involving the recurrent and pathogenic variant p.(Arg65*) associated with heterozygous missenses variants in trans were identified in both cases. Two unreported missense variants were detected: p.(Arg102Trp) (P1) and p.(Arg135Trp) (P2) both located in the NAD-binding domain of the protein. Missense variants p.(Arg102Trp) and p.(Arg135Trp) involved highly conserved residues with CADD scores of 26.0 and 24.5, and were respectively predicted to bound C-terminal aspartates in structural modelizations.
In conclusion, we identified two new and close pathogenic arginine to tryptophane missenses in the same domain and in trans of a known pathogenic nonsense, expanding the phenotype of *UGDH* syndrome to a broader presentation from early infantile intractable encephalopathy to severe ID without epilepsy.

PrgmNr 2201 - The Genetic Architecture of Neurological Diseases in Khyber Pakhtunkhwa-Pakistan

View session detail

Author Block: **M. Ilyas**[1], H. Houlden[2]; [1]Ctr. for Omic Sci., Islamia Coll. Univ., Peshawar, Pakistan, [2]Inst. of Neurology, Univ. Coll. London, London, United Kingdom
Disclosure Block: **M. Ilyas:** None.
Neuro-developmental disorders have become a major public health problem in Pakistan in recent years. Its treatment and management strategies are a daunting problem in Pakistan due to lack of funding, poorly developed primary and basic health care facilities and weak political processes. This increasing rate depends upon certain rare, neglected and often difficult to understand neurological conditions. These disorders therefore require special attention, particularly in rural areas where these disorders remains unexplored. Identifying specific genetic markers may provide a useful explanation for disease etiology, molecular characterization and pathogenesis. We used Next generation sequencing to investigate the novel causative gene(s) in hundreds of families and understand the role in order to develop effective treatments for neurological and neuro-developmental disorders. Clinical examination was performed on more than 50 consanguineous families with affected children. Using exome sequencing, the causative genetic variant was clarified in the families, and 25 convincing candidate genes were identified. It is further expected that personal genetic profiling will also become more relevant, with implications for patient care in line with the proposed idea of personalized medicine.

PrgmNr 2202 - A large deletion in *EVER1* gene: Identification and validation through amplicon-based Next Generation Sequencing

View session detail

Author Block: **A. Godfred**, Z. Thomas, L. Ravichandran, A. Joseph, D. Peter, A. A. George, S. A. Pulimood, P. Gaikwad, T. V. Paul, N. Thomas, A. Chapla; Christian Med. Coll., Vellore, India
Disclosure Block: **A. Godfred:** None.
Mutations in *EVER1* and *EVER2* genes are associated with Epidermodysplasia verruciformis (EV), a rare autosomal recessive genodermatosis. Mutation screening in these genes was carried out in an Indian family with clinically diagnosed EV using amplicon-based Next Generation Sequencing (NGS). Novel primer sets were designed to amplify the coding and splice site regions in these genes in healthy controls and two affected probands. This revealed no amplification with primer set for exons 16 to 18 in the *EVER1* gene of both the probands. To evaluate the possibility of a deletion at the site of the missing amplicon, we designed a Long-range PCR spanning the length of exon 15-20 and NGS was carried out to delineate the deleted segment. This Long range amplicon in the probands and controls was sheared, adapter ligated and size selected, followed by clonal amplification with emulsion PCR and sequencing on Ion Torrent PGM. The generated BAM files were analyzed on Integrated Genome Viewer (IGV), showing a homozygous deletion of 2081bp in exon 17 and 18 of this region. The results were also validated with Sanger sequencing.
Screening the family revealed the same homozygous deletion (similar to index cases) in 2 other affected siblings. The parents and two asymptomatic siblings were heterozygous carriers for the deletion, while one healthy sibling was negative. This deletion *EVER1:c.2072_2278*del produces a frameshift and a resultant loss of 68 amino acids in the transmembrane domain of the protein. The truncated protein product then causes a subsequent loss-of-function which can affect zinc homeostasis. Similar clinical presentations associated with several other mutations in the same gene have been described in the past. This mutation is classified as Pathogenic based on ACMG 2015 guidelines but requires further functional validation to confirm its pathogenicity.
These results indicate that a stepwise amplicon-based target enrichment would enable the identification of a homozygous large deletion before sequencing itself. Also, with long-range PCR options, these results can be validated and sequenced with NGS. Utilizing this strategy, we were able to differentiate carriers with this deletion in a heterozygous state from healthy normal. The data obtained now offers an opportunity for genetic counseling for the family and the possibility of early genetic diagnosis and screening of concerned family members. To the best of our knowledge, this is the first large deletion in *EVER1* gene to be identified in an Indian family.

PrgmNr 2203 - A novel deletion causes SLC25A42-associated mitochondrial encephalomyopathy in a Saudi patient: Report of additional founder cases and functional characterization study

View session detail

Author Block: **M. Aldosary**; KFSH&RC, Riyadh, Saudi Arabia
Disclosure Block: **M. Aldosary:** None.

SLC25A42-associated mitochondrial encephalomyopathy is recently described disorder featuring various symptoms such as hypotonia, intellectual disability and developmental delay, epilepsy and, mitochondrial encephalomyopathy. To date, 15 subjects are reported due to one of two, bi-allelic homozygous missense variants and interestingly most of these patients (n=14) are of Saudi origin harboring the same founder variant, c.871A>G:p.Asn291Asp. The remaining affected individuals, a german patient, was identified with a variant at canonical splice site, c.380+2T>A. Here in this study we employed autozygosity detection using affymetrix axiom human mapping assays, whole exome sequencing and confirmatory sanger sequencing on Saudi patients and identified additional six Saudi patients with the disease from four unrelated consanguineous families. While five patients have the very same Saudi founder p.Asn291Asp variant, one subject has a novel deletion. Functional analyses on fibroblasts obtained from this patient revealed that the deletion causes significant decrease in mitochondrial oxygen consumption and ATP production compared to healthy individuals. Moreover, extracellular acidification rate revealed significantly reduced glycolysis, glycolytic capacity, and glycolytic reserve as compared to control individuals. There were no changes in the mitochondrial DNA (mtDNA) content of patient fibroblasts. Immunoblotting experiments revealed significantly diminished protein expression due to the deletion. In summary, our study expands the molecular spectrum of this condition and provides further evidence of mitochondrial dysfunction as a central cause of the disease pathology. This disorder should be included in the differential diagnosis of any patient with unexplained motor and speech delay, recurrent encephalopathy and metabolic acidosis, intermittent or persistent dystonia, lactic acidosis, basal ganglia lesions and Arab ethnicity. Finally, deep brain stimulation should be considered in the management of patients with life altering dystonia.

PrgmNr 2204 - An *MRM2* splicing variant in a complex dystonic syndrome: a second report and incomplete penetrance

View session detail

Author Block: **A. Shafique**[1], B. Arif[1], S. Naz[1], N. Sobreira[2], E. Wohler[2], C. Klein[3], K. Lohmann[3]; [1]Univ. of the Punjab, Lahore, Pakistan, [2]Johns Hopkins Univ., Baltimore, MD, [3]Univ. of LÃ¼ebeck, LÃ¼ebeck, Germany

Disclosure Block: **A. Shafique:** None.

Dystonia involves repetitive movements and muscle contractions leading to abnormal postures. Pathogenic variants of many mitochondrial and nuclear encoded genes cause isolated dystonia or dystonic syndromes. We analyzed a complex dystonic syndrome in a consanguineous family with four affected individuals. The patients and unaffected individuals were investigated by neurologists and doctors at a local hospital. Magnetic Resonance imaging (MRI), Electromyogram (EMG), laboratory tests including ceruloplasmin and copper levels were completed. The dystonic phenotype manifested in all patients in early childhood which worsened with time. Symptoms started in the upper limbs, and progressively involved the lower limbs and the trunk. The patients were wheel chair bound by the age of 10 years. MRI, EMG and other laboratory tests were normal. Genome-wide mapping was performed on DNA of all participants using GeneChipÂ® Human Mapping 250K Nsp Array. No significant homozygous region which was only present in the affected individuals with perfect segregation was found. An 8.1cM region on chromosome 7p22.3-22.1 was identified which was heterozygous in the obligate carriers and homozygous in all affected individuals as well as in their unaffected 25-year old sibling. Exome sequencing was carried out for two patients. A homozygous splice site variant in *MRM2*, located within the candidate region on chromosome 7, was identified. The variant affected all *MRM2* isoforms, was extremely rare (gnomAD allele frequency 0.00002-no homozygotes) and absent in 150 ethnically matched unaffected controls. No other variant was identified which could explain the phenotype of the patients. Sanger sequencing on samples of all participants confirmed that, similar to the linkage results, the variant in *MRM2* was heterozygous in the obligate carriers and homozygous in the samples of the affected individuals and the unaffected older sibling. The *MRM2* encoded methyltransferase catalyzes the formation of 2'-O-methyluridine within the 16S mitochondrial ribosomal RNA. Previously, a homozygous deleterious missense variant c.567G>A;p.(Gly189Arg) in *MRM2* was linked to Mitochondrial DNA depletion syndrome 17 (MIM#618567). The affected child had a complex movement disorder with multi-organ failure and died in early childhood. Muscular weakness, involuntary and repetitive movements were shared symptoms in affected individuals in our family and the reported case. Our work extends the allelic spectrum of *MRM2* variants including reduced penetrance in the unaffected sibling. This study was supported by the BHCMG, USA, the DFG, & HEC, Pakistan.

PrgmNr 2205 - Associate professor of clinical genetics

View session detail

Author Block: **G. A. Otaify**[1], M. S. Abdel-Hamid[2], S. A. Temtamy[3], M. S. Aglan[4]; [1]Clinical Genetics Dept., Human Genetics and Genome Res. Div., Natl. Res. Ctr., Cairo, Egypt, [2]Med. Molecular Genetics Dept., Human Genetics and Genome Res. Div., Natl. Res. Ctr., Cairo, Egypt, [3]Natl. Res Ctr, Cairo, Egypt, [4]Natl. Res. Ctr., Cairo, Egypt

Disclosure Block: **G.A. Otaify:** None.

Five novel mutations in *FKBP10* and *PLOD2* expanding the molecular and phenotypic spectrum in 10 new patients with rare Bruck syndrome

Bruck Syndrome (BS) is an autosomal recessive disorder characterized by osteogenesis imperfecta (OI) associated with congenital contractures and caused by mutations in *FKBP10* and *PLOD2* genes (BS1 and BS2 respectively). *FKBP10* mutations were found also to cause moderately severe recessive OI type XI. BS is very rare with only 49 cases reported in literature from various ethnic groups. Here we describe 11 new patients from 8 unrelated consanguineous Egyptian families with BS. All patients had white sclerae, recurrent fractures, multiple joint contractures, scoliosis and osteoporosis with variable range of severity. Mutational analysis of *FKBP10* and *PLOD2* genes revealed seven pathogenic variants including 5 novel mutations. *FKBP10* mutations were more common and found in 5 families (62.5 %) while *PLOD2* mutations were identified in 3 families (37.5 %). All *FKBP10* mutations were protein truncating except for one new missense mutation (c.698C>T, p.P233L), while the three *PLOD2* mutations were missense including the novel one. All mutations were homozygous in the probands and affected sibs and heterozygous in their respective parents with no clear genotype-phenotype correlation. Interestingly, the two sibs with the *FKBP10* missense mutation had variable phenotypic presentation. The older sister had joint contractures, severe kyphoscoliosis, hydrocephalus and osteoporosis while the younger brother had multiple fractures, severe osteoporosis and no contractures but rather joint laxity. In conclusion, this study describes five novel mutations in *FKBP10* and *PLOD2* thus expanding the mutational spectrum of the rare BS. In addition, our results further confirm the phenotypic overlap between OI and Bruck syndrome raising the concern of considering BS as a variant of OI and not a separate syndrome.

PrgmNr 2206 - Bi-allelic missense variant p.Thr368Ala in DLST is associated with 2-oxoglutarate dehydrogenase complex deficiency related neurometabolic disorder

View session detail

Author Block: **P. Upadhyai**[1], M. do Rosario[2], P. H. Somashekar[1], K. M. Girisha[3], A. Shukla[4]; [1]Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, [2]Kasturba Med. Coll., Manipal, MAHE, Manipal, India, Manipal, India, [3]Kasturba Med. Coll., Manipal, Manipal, India, [4]Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, Manipal, Udupi, India
Disclosure Block: **P. Upadhyai:** None.
Introduction: The human 2-oxoglutarate dehydrogenase complex (OGDHC) is a multi-subunit mitochondrial tricarboxylic acid (TCA) cycle enzyme that catalyses the decarboxylation of alpha ketoglutarate to succinyl-CoA, reducing NAD+ to NADH. It is composed of multiple copies of three components: 2-oxoglutarate dehydrogenase (OGDH), dihydrolipoyl succinyltransferase (DLST) and dihydrolipoamide dehydrogenase (DLD). Bi-allelic variants in DLD cause dihydrolipoamide dehydrogenase deficiency (MIM #246900) and defect in OGDH has been proposed to cause alpha-ketoglutarate dehydrogenase deficiency (MIM %203740). **Methods:** Clinical examination and magnetic resonance imaging was performed in the proband followed by whole exome sequencing (WES). Segregation analysis was performed by Sanger sequencing. Metabolic investigations were performed using gas chromatography mass spectrometry in blood and urine from the subject. Molecular evaluation of transcript processing, levels and protein expression were carried out by reverse transcription PCR (RT-PCR), reverse transcription quantitative real-time PCR (RT-qPCR) and immunoblotting, respectively using subject fibroblast cells. **Results:** We ascertained a nine-month old female child born with global developmental delay, microcephaly, episodic decompensation, persistent metabolic and lactic acidosis, hyperammonemia and hyperketonemia. Brain imaging revealed atrophy of caudate and putamen. Her clinical features overlap with those reported previously in OGDH and DLD deficiency. WES revealed a missense variant, c.1102A>G (p.Thr368Ala) in DLST (NM_001933.5) that was present in homozygous state in the subject and her parents were heterozygous carriers. The mutation occurs at a highly conserved residue located within the C-terminal acyltransferase inner core or catalytic domain (ICD) of DLST. *In silico* mutagenesis predicted that the mutant residue caused loss of interaction with Asn299 and abrogation of one of two polar contacts with Asp316, as well as diminished mutant protein stability. *DLST* mRNA levels and processing were unchanged. However, DLST protein levels appeared to be significantly reduced. Further congruent with impaired OGDHC enzymatic activity, the levels of alpha ketoglutaric acid and its derivative 2 hydroxyglutaric acid were also elevated in her. **Conclusion:** Taken together we propose DLST as a novel candidate gene for OGDHC deficiency related neurometabolic disorder. This would be further substantiated with future reports of additional individuals and extended functional studies in cells and model organisms.

PrgmNr 2207 - Drug repurposing based on multi-omics data for osteoporosis

View session detail

Author Block: **D-Y. Liu**[1], S. Liang[2], Y. Zhang[1], J-C. Liu[2], Y. Liu[1], H-X. Zhang[1], X-H. Meng[2], H-M. Xiao[2], L-J. Tan[1], H-W. Deng[3]; [1]Lab. of Molecular and Statistical Genetics, Coll. of Life Sci., Hunan Normal Univ., Changsha, China, [2]Ctr. for System Biology, Data Sci., and Reproductive Hlth., Sch. of Basic Med. Sci., Central South Univ., Changsha, China, [3]Tulane Univ., New Orleans, LA

Disclosure Block: **D. Liu:** None.

Osteoporosis is a worldwide public health problem characterized by low bone mineral density (BMD) and a high risk of osteoporotic fracture. The currently available drugs for osteoporosis provide symptomatic benefits, but there is no desirable drug to cure the disease. Repurposing of drugs is a widely used approach to investigate the efficacy of approved drugs for a new indication because of lower overall development costs and shorter development timelines. In the present study, we used driver signaling network identification (DSNI) and drug functional network (DFN) based drug repurposing methods to recommend drug candidates for the treatment of osteoporosis. Osteoporosis driver signaling networks (ODSN, including 1039 genes and 521 subnetworks) were established based on multi-omics data, and drugs functional networks (including 196 drugs and 29 drug functional modules) from the Library of Integrated Network-based Cellular Signatures (LINCS) were determined based on drug similarity. By integrating ODSN and DFN with transcriptome data from LINCS and calculating target effect scores, we found that 8 drugs (including Acebutolol, Amoxapine, Acenocoumarol, Artesunate, Armodafinil, Abacavir, Aminocaproic-acid, and Amiloride.) showed top-ranked scores to display potential for osteoporosis treatment. Furthermore, Mendelian Randomization analysis was used to determine the positive or negative effect on BMD using drug candidates targets cis-eQTL (expression Quantitative Trait Loci) and BMD GWAS data. We found that Acebutolol through inhibiting expression of target ADRB2 (effect = -0.05, P-value = 1.06E-3) and Amiloride through inhibiting expression of target AOC1 (effect = -0.02, P-value = 2.03E-7) showed significant positive association with BMD, while Acebutolol has not previously been proposed as possible treatments for osteoporosis. Besides, Acenocoumarol through inhibiting expression of VKORC1 (effect = 0.01, P-value = 2.92E-48), Aminocaproic-acid through inhibiting expression of PLAT (effect = 0.02, P-value = 3.38E-7), and Armodafinil through inhibiting expression of target SLC6A3 (effect = 0.01, P-value = 1.40E-4), all decreased BMD, which implied that the risks around these drug use for osteoporosis need to be concerned. These findings investigated potential drug candidates in osteoporosis treatment and further warrant bone loss pharmacovigilance with drugs' long-term use.

PrgmNr 2208 - Exome sequencing in 13 families with monogenic forms of rickets

View session detail

Author Block: **K. M. Girisha**[1], P. Jacob[2], G. S. Bhavani[3], P. Udupa[2], Z. Wang[4], V. H. Sankar[5], K. Delampady[6], A. Dalal[7], N. Kamath[8], S. Ikegawa[4,1], R. D. Shenoy[9]; [1]Kasturba Med. Coll., Manipal, India, [2]Kasturba Med. Coll., Manipal, Manipal, India, [3]Manipal Academy of Higher Ed., Manipal, India, [4]Lab. for Bone and Joint Diseases, RIKEN Ctr. for Integrative Med. Sci., Tokyo, Japan, [5]Genetics Clinic, Dept. of Pediatrics, SAT Hosp., Government Med. Coll., Thiruvananthapuram, India, [6]Dept. of Endocrinology, AJ Hosp. & Res. Ctr., Mangalore, India, [7]Diagnostics Div., Ctr. for DNA Fingerprinting & Diagnostics, Hyderabad, India, [8]Kasturba Med. Coll., Manipal Univ., Mangalore, India, [9]Dept. of Pediatrics, K S Hegde Med. Academy, Nitte Univ., Mangalore, India
Disclosure Block: **K.M. Girisha:** None.
Introduction Rickets is a disorder of growing bone caused by defects in the metabolism of calcium, phosphate and/or vitamin D or from nutritional deficiency of vitamin D. Hereditary forms are very rare, and account for about 13% patients with rickets. Eighteen genes are associated with monogenic forms of rickets. **Methods** Sixteen individuals from 13 families were recruited for the study. Detailed clinical features, radiological findings and family history were recorded. Calcium and vitamin D metabolism was assessed by suitable tests. Exome sequencing was performed in the probands from 13 families. The data were analyzed and the rare variants in genes likely to cause the disease were prioritized using standard protocols. **Results** This cohort consists of 11 females and five males with rickets in the age range of 2-21 years. Molecular diagnoses were established for 10 families: vitamin-D dependent rickets type I due to biallelic pathogenic variants in *CYP27B1* (three families); vitamin D resistant rickets, type IIA due to biallelic pathogenic variants in *VDR* (one family); vitamin-D dependent rickets type IB due to pathogenic compound heterozygous variants in *CYP2R1* (one family); hypophosphatemic rickets in four families (pathogenic variants in *PHEX* causing X-linked dominant hypophosphatemic rickets in three and a homozygous pathogenic variant in *SLC34A3* causing hypophosphatemic rickets with hypercalciuria in a family), and biallelic pathogenic variant in *SLC2A2* causing Fanconi-Bickel syndrome. A total of 12 disease-causing variants were identified, of which six were novel. Three families had no diagnosis after exome sequencing in two sibs each. **Conclusion** Our study helped in elucidation of 12 disease-causing variants in six genes associated with monogenic form of rickets in 10 families. Failure of molecular diagnoses in three families even after exome sequencing suggests yet unknown genetic etiology of these disorders.

PrgmNr 2209 - Further delineation of *RPL13*-related skeletal dysplasia

View session detail

Author Block: **P. Jacob**[1], P. Udupa[1], G. S. Bhavani[2], A. B. Dalal[3], K. M. Girisha[4]; [1]Kasturba Med. Coll., Manipal, Manipal Academy of Higher Ed., Manipal, India, Udupi, India, [2]Kasturba Med. Coll., Manipal, India, [3]Ctr. for DNA Fingerprinting and Diagnostics, Hyderabad, India, [4]Kasturba Med. Coll., Manipal, Manipal, India

Disclosure Block: **P. Jacob:** None.

Introduction Spondyloepimetaphyseal dysplasia, Isidor-Toutain type (MIM*618728) is a recently elucidated skeletal dysplasia caused by heterozygous variants in *RPL13* (ribosomal protein, required for pre-ribosomal RNA processing). Till date, nine patients have been reported in the literature. **Methods** Two unrelated families were ascertained. Nine patients from two families were ascertained (Family 1 had eight affected members). Detailed clinical features, radiological findings and family history were recorded. Exome sequencing was performed in two affected individuals of family 1 and in the proband of family 2. Data analysis and interpretation was performed using standard procedures. **Results** Family 1 consists of eight affected members with variable levels of phenotypic severity. The proband at age five years had severe short stature, genu varum, wide wrists, delayed carpal bone ossification and joint laxity. Radiographic features included platyspondyly, bowed femora, severe epiphyseal and metaphyseal dysplasia and exaggerated lumbar lordosis. The other family members demonstrated variable clinical and radiological phenotypes including non-penetrance in one individual. The proband from family 2 at age nine years demonstrated short stature, short trunk and mild pectus excavatum, delayed carpal bone ossification, small and irregular epiphysis of radius and ulna, irregular vertebral end plates and sclerotic changes in the metaphysis of proximal femoral metaphysis. Magnetic resonance imaging showed widening of femoral neck with bilateral symmetrical striated area in the metaphyseal region of proximal femur, osteonecrosis of left femoral head and slipped capital femoral epiphysis on right side. Exome sequencing revealed, a pathogenic variant, c.548G>A p.(Arg183His) in heterozygous state in *RPL13* known to cause Spondyloepimetaphyseal dysplasia, Isidor-Toutain type in both the families. This variant was also present in an unaffected member of family 1 suggesting non-penetrance. A novel missense change at the same amino acid residue (p.Arg183Pro) has been reported recently. **Conclusion** We report two more families with *RPL13*-related skeletal dysplasia with variable clinical features and report an instance of non-penetrance. With this study, we expand the phenotypic spectrum of this rare disease entity.

PrgmNr 2210 - Genetic linkage analysis identifies nuclear regions that modify the phenotype of m.3243A>G-related mitochondrial disease

View session detail

Author Block: **R. Boggan**[1], Y. Ng[1], L. S. Taylor[1], I. G. Franklin[1], C. L. Alston[1], E. L. Blakely[1], B. BÃ¼chner[2], E. Bugiardini[3], K. Colclough[4], G. S. Gorman[1], C. Feeney[1], M. G. Hanna[3], A. T. Hattersley[5], T. Klopstock[2,6,7], C. Kornblum[8], M. Mancuso[9], K. A. Patel[5], R. D. S. Pitceathly[3], C. Pizzamiglio[3], H. Prokisch[10,11], J. SchÃ¤fer[12], A. M. Schaefer[1], M. H. Shepherd[5], A. Thaele[13], D. M. Turnbull[1], C. E. Woodward[14], R. McFarland[1], R. W. Taylor[1], H. J. Cordell[15], S. J. Pickett[1]; [1]Wellcome Ctr. for Mitochondrial Res. and Inst. for Translational and Clinical Res., Newcastle upon Tyne, United Kingdom, [2]Dept. of Neurology, Friedrich-Baur-Inst., Univ. Hosp. of the Ludwig-Maximilians-Univ. (LMU Klinikum), Munich, Germany, [3]Dept. of Neuromuscular Diseases, UCL Queen Square Inst. of Neurology and The Natl. Hosp. for Neurology and Neurosurgery, London, United Kingdom, [4]Exeter Genomics Lab., Royal Devon and Exeter NHS Fndn. Trust, Exeter, United Kingdom, [5]Inst. of BioMed. and Clinical Sci., Univ. of Exeter Med. Sch., Exeter, United Kingdom, [6]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany, [7]German Ctr. for Neurodegenerative Diseases (DZNE), Munich, Germany, [8]Dept. of Neurology, Univ. Hosp. Bonn, Bonn, Germany, [9]Neurological Inst. of Pisa, Pisa, Italy, [10]Inst. of Human Genetics, Sch. of Med., Technische Univ. MÃ¼nchen, MÃ¼nchen, Germany, [11]Inst. of Neurogenomics, Helmholtz Zentrum MÃ¼nchen - German Res. Ctr. for Environmental Hlth., Neuherberg, Germany, [12]Dept. of Neurology, Univ.sklinikum Carl Gustav Carus, Technische Univ. Dresden, Dresden, Germany, [13]Dept. of Neurology, Martin-Luther-Univ. Halle-Wittenberg, 06120 Halle (Saale), Germany, [14]Neurogenetics Unit, The Natl. Hosp. for Neurology and Neurosurgery, London, United Kingdom, [15]Newcastle Univ., Newcastle upon Tyne, United Kingdom

Disclosure Block: **R. Boggan:** None.

Mitochondrial function is under bigenomic control, meaning pathogenic variants in both the nuclear and mitochondrial genomes (mtDNA) result in clinical mitochondrial disease. The most common pathogenic, heteroplasmic, mtDNA variant m.3243A>G (NC_012920.1), is associated with extensive clinical heterogeneity. Established risk factors (age and m.3243A>G variant level) explain only a small proportion of phenotypic variability, whereas high to moderate estimates of heritability for some m.3243A>G-related phenotypes provide evidence for the influence of unidentified nuclear factors. We aimed to explore the genetic aetiology of this variability and locate these factors using genetic linkage analysis.

We identified 238 deeply-phenotyped individual m.3243A>G carriers within 88 pedigrees recruited from: the UK Mitochondrial Disease Patient Cohort; the German Network for Mitochondrial Diseases (mitoNET); the Nationwide Italian Collaborative Network of Mitochondrial Diseases; and the Exeter Centre of Excellence for Diabetes Research. Using the established risk factors of age and m.3243A>G variant level as predictors, we performed logistic regression to quantify the residual unexplained variation in the risk of developing sixteen recognised m.3243A>G-related phenotypes, defined using the validated Newcastle Mitochondrial Disease Score for Adults. To scan the nuclear genome for regions linked to this residual variation, we performed genome-wide Haseman-Elston regression-based linkage analysis, implemented in Merlin-REGRESS, using ~8000 nuclear SNV markers.

For eight m.3243A>G-related phenotypes (cardiovascular involvement, cognition, dysphonia-dysarthria, gastro-intestinal dysmotility, hearing impairment, migraine, neuropathy and seizures), no regions of interest (LOD>1.8) were identified, indicating that the aetiology of these traits is likely to be highly complex, with no major nuclear loci influencing risk. For seven phenotypes (cerebellar ataxia, chronic progressive external ophthalmoplegia, diabetes, myopathy, psychiatric disturbance, ptosis and stroke-like episodes) we identified at least one region of interest (LOD>1.8). For encephalopathy we detected seven regions of interest, including one (LOD>3.3) on the long arm of

chromosome 5, suggesting that a small number of nuclear genetic factors play a key role in the development of this severe neurological phenotype. Association analysis and whole-genome sequencing studies are being pursued to refine these genomic regions.

PrgmNr 2211 - High prevalence of *DMD* gene mutations in Duchenne muscular dystrophy in South Indian population

View session detail

Author Block: **K. Vaidyanathan**, P. K. Manuprasad, C. K. Reshma, S. Nampoothiri; Amrita Inst. of Med. Sci. & Res. Ctr., Kochi, Kerala, India, Kochi, India
Disclosure Block: **K. Vaidyanathan:** None.
Background Alterations in the DMD gene, encoding the dystrophin protein, are known to be common in Duchenne muscular dystrophy (DMD) and Becker-type dystrophy. The DMD gene is located on the X-chromosome. Many mutations are known to be associated with DMD; the commonest being gene deletions/duplications. The objective of our study was to find the variations in DMD gene in confirmed cases of DMD/Becker-type dystrophies. We have studied DMD gene deletion/duplication by Multiplex Ligation-dependent Probe Amplification (MLPA) technique over a two year period (2018-2019). Methods The subjects in this study presented to the Department of Pediatric Genetics, Amrita Institute of Medical Science, Kochi, India during a two-year period (2018-2019). 48 patients with symptoms suggestive of DMD/Becker-type dystrophy were enrolled into the study. They were studied by the MLPA technique. Results 30 patients were detected to have DMD. The commonest changes detected were exons 44-55 deletions (9 patients). Other changes detected were exons 1-44 deletions (5 patients), duplications (3 patients) and exon 55-79 deletions (1 patient). 12 patients tested negative for DMD gene alterations. Conclusions Alterations in DMD gene were frequently found in patients (18/30, 60%) with DMD/Becker-type dystrophy. The commonest alterations found were deletions in exons 44-55 of the DMD gene.

PrgmNr 2212 - Identification of novel candidate genes and expansion of the phenotypic spectrum in a large skeletal dysplasias cohort

View session detail

Author Block: **D. Uludag Alkaya**[1], K. Bilguvar[2], M. GÃ¼nel[3], B. TÃ¼ysÃ¼z[4]; [1]Dept. of Pediatric Genetics, Istanbul Univ.-Cerrahpasa, Istanbul, Turkey, Istanbul, Turkey, [2]Yale Univ., Dept. of Genetics, New Haven, CT, [3]Yale Univ., Dept. of Neurosurgery, New Haven, CT, [4]Istanbul Univ.-Cerrahpasa, Istanbul, Turkey

Disclosure Block: **D. Uludag Alkaya:** None.

Skeletal dysplasias are a large group of disorders that mostly manifest with disproportionate short stature, in which bone and/or cartilage tissue is affected. To date, more than 450 skeletal dysplasia syndromes have been identified and 437 different genes related to these syndromes have been reported. In patients with rare and/or similar signs of skeletal dysplasia, whole exome sequencing can determine the molecular basis of the present disorder. In this study, we aimed to investigate genes related to skeletal dysplasia in a large cohort of patients using the whole exome sequencing. We focused on patients with skeletal dysplasia that could not be diagnosed clinically or molecularly to identify new candidate genes and to contribute to phenotypic expansion. 102 patients from 78 families were included. Molecular diagnosis was confirmed in 55 families (70.5%). The novel candidate skeletal dysplasia genes, *ADAMTS1, CKAP2, MBOAT1, FBN3* were proposed in four of these families. We expanded the phenotype in very recently discovered skeletal dysplasia syndromes such as *PRKG2, GAD1, RPL13, PRKG2, SLC10A7, SIK3* related disorders. Also, novel findings were described in well-known skeletal dysplasia syndromes such as gonadal mosaicism, recessive inheritance pattern in *COL2A1* related disorders. The advanced diagnostic success rates in our clinic is a result of having access to a large patient cohort, selecting relevant patients carefully for this specific study based on abundant clinical and radiological data and implementing appropriate correlation studies between the clinical data sets and patientsâ genetic information.

PrgmNr 2213 - Long-term prognosis and genetic background of cardiomyopathy in 223 mitochondrial disease children

View session detail

Author Block: **A. Okazaki**[1], A. Matsunaga[2], Y. Yatsuka[1], K. R. Nitta[1], Y. Kishita[1], A. Sugiura[1], Y. Sugiyama[2], T. Fushimi[2], M. Shimura[2], K. Ichimoto[2], M. Tajika[2], T. Ebihara[2], T. Matsuhashi[2], T. Tsuruoka[2], T. Hirata[3], H. Harashima[4], A. Takeda[5], A. Ohtake[6], K. Murayama[2], Y. Okazaki[7]; [1]Juntendo Univ., Bunkyo-ku, Japan, [2]Chiba Children's Hosp., Chiba, Japan, [3]RIKEN, Yokohama, Japan, [4]Saitama Med. Univ., moro, Japan, [5]Hokkaido Univ., Sapporo, Japan, [6]Saitama Med. Univ., Moro, Japan, [7]Juntendo Univ., Bunkyo-ku, Tokyo, Japan

Disclosure Block: **A. Okazaki:** None.

Background: Cardiomyopathy is a risk factor for poor prognosis in children with mitochondrial disease. However, other risk factors including genetic factors associated with poor prognosis in mitochondrial disease has yet to be elucidated fully.**Methods and Results**: Between January 2004 and September 2019, we enrolled 223 consecutive children with mitochondrial disease aged Conclusion: In children with mitochondrial disease, cardiomyopathy was common (22%) and was associated with increased mortality. LV hypertrophy, neonatal onset and chromosomal aberrations were independent predictors of all-cause mortality. Prognosis is particularly unfavourable if LV hypertrophy is combined with neonatal onset and/or chromosomal aberrations.

PrgmNr 2214 - Lysosomal storage disorders at a tertiary care centre in a developing country

View session detail

Author Block: **I. Panigrahi**, A. Angral, P. Kaur; PGIMER, Chandigarh, India
Disclosure Block: **I. Panigrahi:** None.
Introduction: The diagnosis of lysosomal storage disorders (LSDs) can be challenging because o varied clinical presentations at different ages. Gaucher disease is a prototype disorder for which various forms of therapy are available. Enzyme replacement therapy (ERT) is now available here for MPS I, MPS II, Gaucher disease, Fabry disease and Pompe disease. We report the spectrum of LSDs at a tertiary care center**. Methods:** The families of patients of lysosomal storage disorders visiting the Genetic Clinic, or admitted in the Genetic ward in last 12 years were included in this retrospective analysis. Consent for genetic testing or antenatal testing was as per standard protocols and recommendations of the PNDT Act. Mutations were done only in some families, especially to facilitate prenatal diagnosis. In families in which PND was done, the testing was done by enzyme analysis and/or mutation testing by Sanger sequencing on chorionic villus or amniotic fluid samples. **Results:** In last 12 years, 35 families with Gaucher disease, 27 families with MPS, 10 families with Niemann Pick disease (NPD), 8 families of Pompe disease, 5 families with GM1 gangliosidosis, 4 families with GM2 gangliosidosis and 2 families with Fabry disease were seen, evaluated and counselled in the Genetic Clinic or Genetic ward of the Institute. The patients were from adjoining states. The diagnostic delay varied from 2-10 years. PND was done in 3 Gaucher disease, 2 each for GM1 and GM2 gangliosidosis, 2 NPD, 2 MPS III and 2 Pompe disease families. Most patients with Gaucher disease showed the *GBA*: c.1448T>C(L444P) mutation in homozygous or heterozygous state. Therapy in form of ERT was given in 4 patients with Gaucher disease, 2 patients with MPS II, and one patient with Pompe disease. It could not be given in most eligible patients because of financial constraints. .**Conclusions:** LSDs can present with hepatosplenomegaly, cytopenias, renal disease, cardiomyopathy, neurological and skeletal features. In developing countries, usually the diagnosis is delayed, all patients do not have access to therapy, and prevention by prenatal diagnosis is major preventive option in these families. We could diagnose many patients by enzyme testing, and could plan and perform prenatal diagnosis in some of the families by enzyme or molecular diagnosis. In absence of rare disease model, and government support for therapy for many families; the parents run from pillar to post for ensuring therapy for their children. With the National policy for Rare diseases (NPRD) implementation, and streamlining of the services and prioritisation of therapy in eligible patients, the situation is likely to change in next few years.

PrgmNr 2215 - Mechanism of WRN loss for causing short stature in Werner syndrome

View session detail

Author Block: **H. Cheung**, O. Leung, T. Yiu; The Chinese Univ. of Hong Kong, Shatin, Hong Kong
Disclosure Block: **H. Cheung:** None.
Werner Syndrome is a premature aging disease characterized by early onset of many age-related phenotypes in the adulthood. It is a monogenic disease caused by recessive mutations in the WRN gene, a member of the conserved RECQ DNA helicases. The first phenotype of WS is typically displayed in the skeletal system, with approximately 95% of reported WS cases showed short stature. Individuals diagnosed as WS have an abnormally slow growth rate, and growth usually stops at puberty. As a result, affected individuals develop short stature. The mechanism through which WRN loss causes short stature, and the underlying genetic basis are not clearly elucidated. To determine the cause of short stature in WS, we used human pluripotent stem cell (hPSC) model for identifying deferentially expressed genes in the differentiated mesenchymal cells. Transcriptome analysis revealed that *short stature homeobox* (SHOX) gene was one of the downregulated targets in WS. SHOX is known to play a vital role in chondrogenesis and bone development. As SHOX deficiency is a frequent cause of short stature, we hypothesized that SHOX is downregulated in WRN mutated cells. Knockout of WRN or SHOX in human mesenchymal stem cells and hPSC impaired chondrocyte differentiation, as shown by reduced expression of chondrocyte markers SOX9 and COL2A1. Furthermore, we identified a molecular mechanism that wild-type WRN protein binds to SHOX promoter and resolves the G-quadruplex structure which hinders the transcription of SHOX. The current work suggests a novel mechanism through which SHOX insufficiency in WS is connected with bone development.

PrgmNr 2216 - Molecular diagnostic challenges related to mutations in the low-complexity domain of heterogenous nuclear ribonuclearprotein A1 (HNRNPA1)

View session detail

Author Block: **M. Johari**[1], S. Rusanen[1], M. Savarese[1], P. Jonson[1], J. Sarparanta[1], A. Vihola[1], I. Paramonov[2], L. Matalonga[2], A. TÃ¶pf[3], M. Minot-MyhiÃ©[4], M. Fradin[4], F. Letournel[5], P. Marcorelles[6], A. Nadaj[7], M. Spinazzi[8], P. Hackman[1], B. Udd[1]; [1]FolkhÃ¤lsan Res. Ctr., Univ. of Helsinki, Helsinki, Finland, [2]CNAG-CRG, Ctr. for Genomic Regulation (CRG), The Barcelona Inst. of Sci. and Technology, Barcelona, Spain, [3]The John Walton Muscular Dystrophy Res. Ctr., Translational and Clinical Res. Inst., Newcastle Univ. and Newcastle Hosp. NHS Fndn. Trust, Newcastle upon Tyne, United Kingdom, [4]Service de Genetique, CHU Rennes, Ctr. de Competence des Maladies Neuromusculaires de Rennes, Rennes, France, [5]Lab. of neuropathology, Dept. of pathology, Univ. Hosp. of Angers, Angers, France, [6]Ctr. Hosp.ier Regional Univ.ire (CHRU), Brest, France, [7]Neuromuscular Referential Ctr. Nord/Est/Ile-de-France Neurology Dept.. Univ. Hosp. of Strasbourg, Strasbourg, France, [8]Neuromuscular Referral Ctr., Neurology Dept., Univ. Hosp. of Angers, France, Angers, France

Disclosure Block: **M. Johari:** None.

Heterogeneous nuclear ribonucleoproteins are RNA-binding proteins, many of which contain low complexity sequences known as â□□prion-like domainsâ□□ or PrLDs. Mutations in these PrLDs are associated with several diseases such as Multisystem proteinopathies (MSP) and Amyotrophic lateral sclerosis (ALS). These mutations affect the abnormal stress granule (SG) formation and clearance resulting in increased toxicity due to aggregated proteins. Understanding how mutations in heterogenous nuclear ribonuclearprotein A1 (*HNRNPA1*) functionally affect the SGs, holds the key in improving our knowledge of the molecular pathomechanisms of the disease.We identified three different mutations in the PrLD of *HNRNPA1* (NM_031157) associated with dominant myopathy phenotypes in three unrelated families (FAM1-3). In FAM1, the previously reported D314N mutation segregated with a proximal rimmed vacuolar myopathy. The finding was initially missed due to coverage issues in exomes and traditional variant calling on captured regions. In FAM2, published earlier with a distinct distal myopathy phenotype (MPD3), we performed linkage analysis, subsequent Sanger sequencing of candidate genes, exome and genome sequencing (GS), which remained inconclusive. Copy Number Variant analysis of newer high depth GS identified a heterozygous 160 bp deletion in exon 10 of *HNRNPA*1 in the linked region, segregating with the phenotype in this family. The finding again asserted the possible coverage issues plaguing this gene. In FAM3 with severe myopathic atrophy in forearms and hands, we identified a *de novo* heterozygous indel in the same exon 10 causing extension of the HNRNPA1 tail by 3 amino acids. We are currently investigating the effects of these mutations on transcript and protein level. We postulate that different mutations in the PrLD of *HNRNPA1* can give rise to different overlapping phenotypes. Our internal analysis of over 400 exomes and genomes suggests, that due to the genomic complexity of this region some sequencing chemistries do not result in high confidence coverage. The coverage analysis of data from 10638 samples in Solve-RD suggests that exome capture kits have difficulty in covering the GC rich regions of *HNRNPA1* which could lead to cases of â□□missed diagnosisâ□□.

PrgmNr 2217 - Mutations in DNA ligase III cause mitochondrial neurogastrointestinal encephalomyopathy

View session detail

Author Block: **M. Taniguchi**[1], E. Bonora[2], S. Chakrabarty[3], M. Tsutsumi[4], F. Bianco[2], G. Kellaris[5], C. Bergamini[6], H. Kurahashi[7], K. Kosaki[8], T. Toda[9], N. Katsanis[10], F. Duijkers[11], R. De Giorgio[12]; [1]Dep. Clin. Genet. Fujita Hlth.Univ. Hosp., Toyoake, Japan, [2]Dept, Med. and Surgical Sci., Univ. of Bologna, Bologna, Italy, [3]Dept. Mol. Genet., Erasmus MC, Rotterdam, Netherlands, [4]Fujita Hlth.Univ., Toyoake, Japan, [5]Ctr. for Human Disease Modeling, Duke Univ., Durham, NC, [6]Dept. Pharmacy and Biotechnology, Univ. of Bologna, Bologna, Italy, [7]ICMS, Fujita Hlth.Univ., Toyoake, Japan, [8]Keio Univ. Sch. of Med., Tokyo, Japan, [9]The Univ. of Tokyo, Tokyo, Japan, [10]Duke Univ., Durham, NC, [11]AMC, Utrecht, Netherlands, [12]Univ. of Ferrara, Cona, Italy
Disclosure Block: **M. Taniguchi:** None.
Abnormal gut motility is a feature of several mitochondrial encephalomyopathies, and mutations in genes such as *TYMP* and *POLG*, have been linked to these rare diseases. The human genome encodes three DNA ligases, of which only one, ligase III (*LIG3*), has a mitochondrial splice variant and is crucial for mitochondrial health. We investigated the effect of reduced LIG3 activity and resulting mitochondrial dysfunction in seven patients from three independent families, who showed the common occurrence of gut dysmotility and neurological manifestations reminiscent of mitochondrial neurogastrointestinal encephalomyopathy. DNA from these patients was subjected to whole exome sequencing. In all patients, compound heterozygous variants in a new disease gene, *LIG3*, were identified. All variants were predicted to have a damaging effect on the protein. The *LIG3* gene encodes the only mitochondrial DNA (mtDNA) ligase and therefore plays a pivotal role in mtDNA repair and replication. *In vitro* assays in patient-derived cells showed a decrease in LIG3 protein levels and ligase activity. We demonstrated that the *LIG3* gene defects affect mtDNA maintenance, leading to mtDNA depletion without the accumulation of multiple deletions as observed in other mitochondrial disorders. This mitochondrial dysfunction is likely to cause the phenotypes observed in these patients. The most prominent and consistent clinical signs were severe gut dysmotility and neurological abnormalities, including leukoencephalopathy, epilepsy, migraine, stroke-like episodes, and neurogenic bladder. A decrease in the number of myenteric neurons, and increased fibrosis and elastin levels were the most prominent changes in the gut. Cytochrome *c* oxidase (COX) deficient fibres in skeletal muscle were also observed. Disruption of *lig3* in zebrafish reproduced the brain alterations and impaired gut transit *in vivo*. In conclusion, we identified variants in the *LIG3* gene that result in a mitochondrial disease characterized by predominant gut dysmotility, encephalopathy, and neuromuscular abnormalities.

PrgmNr 2218 - NTBC efficacy in alkaptonuria

View session detail

Author Block: **H. Alsharhan**[1,2], H. Alharbi[2], S. M. Kallish[3], J. Sharer[4], S. R. Master[2,3], M. Yudkoff[2,3], M. He[5]; [1]Dept. of Pediatrics, Faculty of Med., Kuwait Univ., Kuwait City, Kuwait, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [4]Univ. of Alabama at Birmingham, Birmingham, AL, [5]Childrenâ□□s Hosp. of philadelphia, Philadelphia, PA
Disclosure Block: **H. Alsharhan:** None.
Alkaptonuria (AKU) is a rare autosomal recessive disorder in tyrosine degradation pathway, characterized by ochronosis, arthritis and significant urinary homogentisic acid (HGA) excretion. There are no preventive or curative treatment options currently available. Nitisinone (NTBC), a drug that is currently approved for treating AKU by the European Medicines Agency but not by the US Food and Drug Administration (FDA), was found to significantly decrease plasma and urine HGA and benzoquinone acetic acid levels while moderately increasing plasma tyrosine. Individuals with hereditary tyrosinemia type I require immediate treatment with NTBC to avoid acute hepatic failure. This is not the case in patients with AKU, in whom clinicians must balance an anticipated reduction of HGA production against the risk of an elevated tyrosine level. We present here three cases with AKU (one child and two adults) who were managed with NTBC. In the first case, a 47-year-old female, treatment (0.3 mg/kg/day, or a total of 20 mg/day) has normalized HGA excretion but caused a rise in blood tyrosine to 521.2 Âµmol/L (normal: 22 - 102 Âµmol/L). Lowering the dose by half maintained negligible excretion of HGA but plasma tyrosine remained elevated at 567 Âµmol/L. Our second case, a 5-year-old boy who received a low dose of NTBC (0.1 mg/kg/day), manifested continued diminished excretion of HGA (101 mg/gm creatinine), but he developed elevated plasma tyrosine (713.1 Âµmol/L) and corneal crystals after 10 months of treatment. The NTBC was discontinued. Similarly, the third patient, a 70-year-old male, developed blepharitis on a daily NTBC dose of 0.03 mg/kg/day (2 mg/day) with plasma tyrosine at 976 Âµmol/L and 86% reduction in HGA to 181 mg/g creatinine. Reduction of NTBC administration to an every-other-day schedule resulted in continued increase of plasma tyrosine (740-799 Âµmol/L) and a 56% reduction of urinary HGA to 540 mg/g creatinine. Our report indicates that the effect of NTBC in lowering HGA is variable among patients and that even seemingly low drug doses can cause undue increases of plasma tyrosine. Careful monitoring of both plasma tyrosine and urinary HGA are essential in order maximally to inhibit production of HGA without elevating systemic tyrosine concentrations to a toxic level that causes serious ocular side effects. Indeed, even ultra-low doses may favor deposition of tyrosine crystals in the cornea. A need for regular monitoring of blood tyrosine is clear. It may be prudent to defer initiation of treatment in young children, given a theoretical concern for the impact of hypertyrosinemia on cognitive development and our observation of corneal crystal deposition.

PrgmNr 2219 - Outcome of Clinical Genetic Testing in Patients with Features Suggestive for Ehlers Danlos Syndrome

View session detail

Author Block: **N. Damseh**[1], L. Dupuis[2], C. Oâ□□Connor[3], r. Oh[3], D. J. Stavropoulos[4], S. Schwartz[3], R. Mendoza-Londono[2]; [1]Clinical and Metabolic Genetics, Hosp. for Sick Children, TORONTO, ON, Canada, [2]Hosp. for Sick Children, Toronto, ON, Canada, [3]The Hosp. for Sick Children, TORONTO, ON, Canada, [4]Hosp for Sick Children, Toronto, ON, Canada

Disclosure Block: **N. Damseh:** None.

Ehlers Danlos Syndrome (EDS) is a heterogeneous group of connective tissue disorders (CTD). The cardinal features of EDS are hyperextensible skin, hypermobile joints, easy bruisability, and fragility of the connective tissues leading to a wide variety of clinical manifestations. The EDS diagnosis is usually based on clinical assessment and phenotypeâ□□guided genetic testing. All the EDS subtypes can be confirmed by genetic testing except for hypermobile EDS (hEDS). In this retrospective study, we reviewed 72 patients suspected to have EDS who underwent genetic testing. Eighteen patients (25%) met the clinical criteria for one of the EDS subtypes (other than hEDS), 15 patients (83%) of which received confirmatory molecular genetic diagnoses. Fifty-four patients (75%) didnâ□□t meet full clinical criteria however 13 of these patients (26%) also received a confirmatory molecular genetic diagnosis Different molecular genetic tests were performed (CTD panel n=44, microarray n=23, whole exome sequencing (WES) n=9, targeted gene testing n=4, familial variant testing n=12, other genetic panels n=3). The diagnostic outcome of the CTD genetic panel was 41% (82% in patients who met one of the EDS subtype diagnostic criteria and 27% of patients who did not) while WES had a diagnostic outcome of 22%. We observed that generalized joint hypermobility, poor healing, easy bruising, atrophic scars, skin hyperextensibility, and developmental dysplasia of the hip correlate with positive genetic results. This study explores the utility of 2017 EDS classification criteria and molecular genetic testing in establishing an EDS diagnosis in pediatric population. It also provides some guidance in using molecular genetic testing in children presenting with EDS characteristics.

PrgmNr 2220 - Phenotypic Spectrumof DihydrolipoamideDehydrogenase Deficiencyin Saudi Arabia

View session detail

Author Block: **A. Alfarsi**[1], M. Alfadhel[2], F. Ababna[3], S. Alameer[4], A. Alhashem[5], I. Tabarki[6], A. Alfares[7], F. Al Mutairi[8]; [1]King Abdulaziz Univ. Hosp., Jeddah, Saudi Arabia, [2]King Abdullah Intl. Med. Res. Ctr., King Saud bin Abdulaziz Univ. for Hlth.Sci.. Dept. of Genetics, King Abdulaziz Med. City, Ministry of Natl. Guard-Hlth.Affairs (NGHA), Riyadh, Saudi Arabia, Riyadh, Saudi Arabia, [3]King Abdulaziz Med. City, Riyadh, Saudi Arabia, [4]Dept. of pediatrics, King Abdulaziz Med. City,., Jeddah, Saudi Arabia, [5]Dept. of Pediatrics, Prince Sultan Military Med. City,., Riyadh, Saudi Arabia, [6]Div. of Neurology, Dept. of Pediatrics, Prince Sultan Military Med. City, Riyadh, Saudi Arabia, [7]Dept. of Lab Med., Div. of Translational Pathology, King Abdulaziz Med. City,, Riyadh, Saudi Arabia, [8]KASCH-KAMC, Riyadh, Saudi Arabia

Disclosure Block: **A. Alfarsi:** None.

BACKGROUND: Dihydrolipoamide Dehydrogenase Deficiency is one of the rare metabolic disease with autosomal recessive inheritance. It is clinically heterogeneous with variable presentations, onsets, and biochemical markers. Yet, it difficult to have a genotype-phenotype correlation.
MATERIALS AND METHODS: We retrospectively review the clinical and molecular diagnosis of 8 cases with DLD from four referral center in Riyadh. **RESULTS:** we found hepatic involvement in half of the patient ranging between acute hepatic failure, encephalopathy and chronic hepatitis. One patient had Recurrent episodic hypoglycemia while neurological involvement seen in form of seizure, developmental delay, ataxia, hypotonia and psychomotor symptoms. In regards to biochemical markers, lactic acid, metabolic acidosis and branched chain amino acid are extremely variable across our cohort. 60% of the cohort patients had homozygous variant p.(Gly229Cys) in the DLD gene.
CONCLUSIONS: We describe the largest reported DLD cohort in Saudi Population. The clinical, biochemical, radiological, and molecular characterization have been reviewed. Genotype-phenotype correlation is still unclear, that related to the complexity of the enzyme.

PrgmNr 2221 - The p.L3P (p.Leu3Pro) variant in GLA is not associated with Fabry disease

View session detail

Author Block: **O. Ailioaie**[1], R. MATESAN[2], M. N. KONAN[3], F. JABBOUR[1], J-P. RABES[4], N. MAROUN[2], D. P. GERMAIN[5]; [1]APHP - Paris Saclay Univ., Garches, France, [2]GHT Yvelines Nord, Poissy, France, [3]CHU de Treichville, Abidjan, CÃ´te D'Ivoire, [4]APHP - Paris Saclay Univ., Boulogne, France, [5]Paris Saclay Univ., Univ. of Versailles, Montigny, France
Disclosure Block: **O. Ailioaie:** None.
Background: Fabry disease (FD, OMIM #301500) is an X-linked genetic disease characterized by the deficient activity of the lysosomal alpha-galactosidase. The functional deficiency of the enzyme leads to pathogenic accumulation of Gb3 and lyso-Gb3 responsible for the development of a multi-systemic disease with onset during childhood (acroparesthesia, angiokeratoma) and subsequent complications (hypertrophic cardiomyopathy, chronic kidney disease, stroke) in adulthood leading to reduced life expectancy. Over 1000 variants have been reported in the *GLA* gene. Among those, a significant number are benign or likely benign and not associated with FD. **Patient and Methods**: A 64-year-old male patient of African ancestry with a medical history of familial cardiomyopathy and end-stage renal disease treated by hemodialysis benefited from a molecular screening on dried blood spot (DBS) funded by the pharmaceutical industry that led to the identification a missense variant (p.L3P or p.Leu3Pro) initially reported as of variant of unknown significance (class 3) leading the treating nephrologist to consider the diagnosis of FD on the basis of both clinical and molecular data. The patient was therefore referred to the French Referral Center for Fabry disease for further investigation. **Results:** Interrogation of the ClinVar database gave conflicting results on pathogenicity. Data obtained from GnomAD database showed that the p.L3P variant, while absent in Caucasian and Hispanic subpopulations, was relatively frequent in the African sub-populations (allelic frequency of 0.00319, higher than 0.000265 derived from 620 variants in *GLA*) emphasizing the value of studying multiple ethnic backgrounds in databases. A complete workup was carried out showing normal (100%) alpha-galactosidase A activity, normal lyso-Gb3 values and skin biopsy with no storage when examined with light microscopy allowing us to definitely rule out the diagnosis of Fabry disease. **Discussion:** This case illustrates how surrogate markers may be of considerable help in elucidating the pathogenicity of a molecular variants in some inborn errors of metabolism. Datasharing data with the medical and genetics community is of utmost importance to avoid undue healthcare expenses in individuals bearing benign variants. In the present case, correct assignment of the variant has avoided the very high costs for the healthcare system associated with specific treatment of FD (170,000 USD /year) and the iatrogenic risks associated with intravenous infusions of enzyme therapy. Moreover, it warrants further diagnostic odyssey since the identified variant cannot explain the observed clinical phenotype.

PrgmNr 2222 - Association analysis of rare variants identified risk genes with neonatal respiratory distress syndrome

View session detail

Author Block: **X. Dong**, H. Chen, X. Chen, W. Zhou; Children's Hosp. of Fudan Univ., shanghai, China
Disclosure Block: **X. Dong:** None.
Abstract Background Neonatal respiratory distress syndrome (NRDS) is considered to be primarily caused by deficiency of pulmonary surfactant (PS) in neonates. Genetic abnormalities were recently identified as the important etiologies. **Methods** The newborns were enrolled from the China Neonatal Genomes Project (CNGP, NCT03931707) from August 2016 to June 2020. All patients underwent clinical exome sequencing but no disease-causing genes were found. Rare-variant association analysis were performed to detect the potential genetic risk factors of NRDS. For exploring their functions, a lung-specific gene regulatory network was constructed and the regulatory activity was calculated to study their role in fetal lung developmental. After focusing on genes with continued increasing pattern, a multivariate logistic regression model was used to extract out the independent NRDS contributed risk genes. **Results** A total of 123 RDS and matched 214 control infants with gestational ageâ¥34 weeks were included. Association analyses found 53 high-confidence NRDS risk genes and 10 genes had continued increasing pattern among three stage of normal fetal lung developmental (P=0.019). Among them, 5 were RDS-independent risk genes (*GRN, SFTPB, IL1RN, TJP2, ADAMTS2*) and only *ADAMTS2* showed âhiddenâ feature (with significantly activity pattern change but not expression level change) in regulating lungâs development. Re-analyze the public transcriptomic dataset showed that glucocorticoids could increase the activity of *ADAMTS2* in mouse lung A549 cell lines (PADAMTS2 variants. **Conclusion** NRDS is probably a multigenetic disease with complex genetic risk factors. Except *SFTPB*, we detected other new novel and potential NRDS risk genes, including *ADAMTS2*. Further investigation is necessary to identify their function in the onset of NRDS.

PrgmNr 2223 - Combined immunodeficiency and increased cellular sensitivity to radiation due to a novel DNA Ligase 1 mutation

View session detail

Author Block: **A. Alazami**[1,2], H. Alajlan[1], R. Mohammed[1], H. Al-Mousa[1,2,3]; [1]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia, [2]Saudi Human Genome Program, King Abdulaziz City for Sci. and Technology, Riyadh, Saudi Arabia, [3]Coll. of Med., Alfaisal Univ., Riyadh, Saudi Arabia
Disclosure Block: **A. Alazami:** None.
DNA Ligase 1 (*LIG1*) plays a critical role in the joining of Okazaki fragments during DNA replication, as well as in single- and double-stranded break repair pathways. Only six patients have been reported to date with mutations in this gene, exhibiting a range of phenotypes such as hepatosplenomegaly, sun sensitivity, severe eczema, severe anemia and recurrent infections. Here we report the case of a baby girl who presented with recurrent chest infections, significant erythroid dysplasia, neutropenia and combined immunodeficiency. She underwent successful bone marrow transplantation 3 years ago at the age of 6 months. Whole exome sequencing revealed homozygosity for a novel *LIG1* mutation (A556T), which segregated with the disease state. Immunophenotyping displayed depressed T cell levels along with most CD4+ and CD8+ subsets, but with an increase in CD4+ central memory cells, and with a reduction in the B memory subset. Patient fibroblast cells showed severe reduction in proliferation rates, and exposure to ionizing radiation led to a cell cycle block in the G2/M phase along with significant increase in gamma-H2AX foci as compared to controls. Our findings broaden the phenotypic and allelic heterogeneity associated with this rare autosomal recessive immunodeficiency.

PrgmNr 2224 - Functional characterization of a *Xenopus tropicalis* knockout and a human cellular model of *RCBTB1*-associated inherited retinal disease shows involvement of *RCBTB1* in the cellular response to oxidative stress

View session detail

Author Block: **M. Carron**[1], T. Naert[2], G. Ascari[3], S. Demuynck[2], T. Van Nieuwenhuyzen[2], T. Van Laethem[3], T. Rosseel[3], D. Priem[4], A. Kremer[5], H. Van Landeghem[2], L. Wanneyn[2], S. Hoogstoel[3], F. Coppieters[6], C. GuÃ©rin[7], K. Vleminckx[8], E. De Baere[6]; [1]Dept. of BioMed. Molecular Biology; Ctr. for Med. Genetics, Dept. of Biomolecular Med., Ghent Univ. and Ghent Univ. Hosp.,, Gent, Belgium, [2]Dept. of BioMed. Molecular Biology, Ghent Univ., Gent, Belgium, [3]Ctr. for Med. Genetics, Dept. of Biomolecular Med., Ghent Univ. and Ghent Univ. Hosp., Gent, Belgium, [4]VIB-UGent Ctr. for Inflammation Res., Gent, Belgium, [5]VIB Imaging Core, Gent, Belgium, [6]Ctr. for Med. Genetics, Dept. of Biomolecular Med., Ghent Univ. and Ghent Univ. Hosp., Ghent, Belgium, [7]VIB Bioimaging Core, Gent, Belgium, [8]Dept. of BioMed. Molecular Biology, Ghent Univ., Ghent, Belgium
Disclosure Block: **M. Carron:** None.
We identified *RCBTB1* as a novel disease gene for autosomal recessive inherited retinal disease (IRD), with or without syndromic features, such as primary ovarian insufficiency, goiter, and mild intellectual disability. Patients with biallelic variants in *RCBTB1* display diverse IRD phenotypes such as retinitis pigmentosa, reticular dystrophy and chorioretinal atrophy. Each novel genetic subtype of IRD may shed light on the particularly complex pathways of vision and create opportunities for intervention. So far the function of the ubiquitously expressed *RCBTB1* gene remains unknown. RCBTB1 has been indicated as putative substrate adaptor for the E3 ligase Cullin 3 (CUL3) complex and interactor of UBE2E3, an E2 conjugating enzyme. As both CUL3 and UBE2E3 are associated with regulation of the NRF2/KEAP1/ARE pathway, we hypothesize that *RCBTB1* is involved in NRF2-regulated protection against oxidative stress in the eye, more specifically in the retinal pigment epithelium (RPE). Here, a *Xenopus tropicalis* knockout (KO) animal model was generated using CRISPR/Cas9 gene editing. Histological examination and three-dimensional electron microscopy was performed on retinas of *rcbtb1*$^{-/-}$ frogs. RNA-seq analysis was performed on embryos from the *rcbtb1*$^{-/-}$ animal model treated with $CdCl_2$ and on lymphocytes from cases with biallelic *RCBTB1* variants, treated with H_2O_2. A knockdown (KD) cell line was generated in ARPE-19 cells via viral transduction and in vitro functional assays (flow cytometry, MTT-assay, cell death kinetics) were used to assess the consequences of *RCBTB1* depletion. The *rcbtb1*$^{-/-}$ animals showed changes in the RPE, similar to observations in human cases, including loss of apical-basal cell polarity, loss of cuboidal cell morphology, spreading of the pigment granules and vacuolisation. Via RNA-seq, NRF2 downstream targets and several metallothioneins were found to be differentially expressed, both in the animal and cellular models. The functional assays in ARPE-19 cells revealed that RCBTB1 depletion affects cellular responses to external insults of oxidative stress. We showed that the *Xenopus tropicalis rcbtb1*$^{-/-}$ animal model recapitulates the human IRD phenotype. Both in vivo and in vitro functional data show involvement of *RCBTB1* in the cellular response to oxidative stress. This provides insight into the mechanism underlying *RCBTB1*-associated IRD and uncovers potential future therapeutic opportunities.

PrgmNr 2225 - Genetic and clinical characteristics of inherited retinal dystrophies in Northern Finland

View session detail

Author Block: **L. LÃ¤hteenoja**[1,2], A. Falck[2], E. Rahikkala[1,3]; [1]Dept. of Clinical Genetics, PEDEGO Res. Unit and Med. Res. Ctr. Oulu, Oulu Univ. Hosp. and Univ. of Oulu, Oulu, Finland, [2]Dept. of Ophthalmology, Oulu Univ. Hosp. and Univ. of Oulu, Oulu, Finland, [3]Inst. of Biomedicine, Univ. of Turku, Turku, Finland
Disclosure Block: **L. LÃ¤hteenoja:** None.
Purpose: To describe the genetic and clinical features of patients with genetically confirmed inherited retinal degenerations (IRDs) in the Northern Finnish founder population. **Methods**: A population-based study was conducted in the tertiary catchment area of the Oulu University Hospital between 1996 and 2020. Patients were identified retrospectively by International Classification of Diseases codes in hospital records. The inclusion criterion was genetically confirmed IRD. The pathogenicities of variants were evaluated according to the American College of Medical Genetics and Genomics guidelines (ACMG). The phenotype-genotype correlations were defined. **Results**: 152 patients were diagnosed with a genetically confirmed IRD. The probands were diagnosed most commonly with retinitis pigmentosa, choroideremia, familial exudative vitreoretinopathy or retinoschisis. A total of 225 pathogenic or potentially pathogenic variants in 34 genes were observed, including 98 missense variants, 33 frameshift variants, 41 nonsense variants, 22 canonical splice site variants, 27 deletion variants, three inversion variants, and one duplication variant. The most common genes causing RD were *RPGR*, *CHM*, *FZD4*, *RS1* and *CERKL*. Causal variants were consistent with a recessive mode of inheritance in 49% (33% homozygous, 16% compound heterozygous), dominant in 14%, and X-linked in 35% of cases. 26% had syndromic IRD, including 16 causative genes. 53 patients (35%) with diseases belonging to the Finnish disease heritage were identified. **Conclusion**: IRD is genetically and clinically highly heterogeneous. This study expands the spectrum of disease-causing genes and variants in IRD, which will further facilitate genetic counseling.

PrgmNr 2226 - Genetic variants influencing penetrance to iron overload disorder hereditary haemochromatosis: evidence from UK Biobank

View session detail

Author Block: **L. C. Pilling**, J. L. Atkins, D. Melzer; Univ. of Exeter, Exeter, United Kingdom
Disclosure Block: **L.C. Pilling:** None.

Haemochromatosis is the most common genetic disorder in Northern Europeans, with up to 1 in 150 people homozygous for the *HFE* p.C282Y mutation. p.C282Y increases iron absorption from the diet, resulting in increased morbidity, especially liver disease and cancer, especially in males. Yet penetrance is incomplete and expressivity is heterogeneous, with many homozygous carriers either diagnosed late or not at all, with varying degrees of symptoms. Co-morbidities can include diabetes, arthritis, and osteoporosis, but the factors that modify individual patient trajectories are uncertain. We aimed to identify genetic factors in *HFE* p.C282Y homozygotes affecting likelihood of haemochromatosis diagnosis and co-morbidity status.

In 2,980 *HFE* p.C282Y homozygotes in the UK Biobank cohort we investigated factors that modify penetrance, disease severity, and co-morbidities. Using genetic risk scores (GRS), we found that p.C282Y homozygotes carrying greater numbers of osteoarthritis alleles had increased likelihood of osteoarthritis diagnoses (Odds Ratio per standard deviation of GRS 1.24: 95% Confidence Intervals 1.08 to 1.41, p=0.002), and similarly participants with higher GRS for Rheumatoid arthritis, diabetes, and fracture risk (osteoporosis) had greater risk of the respective co-morbidities. Underlying predisposition to individual diseases may therefore explain part of the phenotypic heterogeneity in haemochromatosis patients.

We also performed genome-wide analysis of haemochromatosis diagnosis in p.C282Y homozygotes and found one locus in the *TMPRSS6* gene (rs9610638) associated with increased likelihood of diagnosis (OR per T allele 1.21: 95% CIs 1.17 to 1.25, $1*10^{-8}$). This locus is known to affect iron levels, including in the brain. In analysis using 11 non-*HFE* iron-associated variants in a GRS, p.C282Y homozygotes with increased iron GRS were substantially more likely to be diagnosed with haemochromatosis (OR for top 20% of GRS 2.36: 95% CIs 1.79 to 3.13, $p=2*10^{-9}$).

Together, these results show that haemochromatosis is a complex disease somewhere between monogenic and polygenic. Patient outcomes could be improved by taking account of more genetic variants, thereby identifying the most at-risk patients.

PrgmNr 2227 - Identification of variants associated with moderate to severe hearing loss

View session detail

Author Block: **S. Naz**[1], M. Ramzan[2], H. Idress[1], R. Faridi[3], Z. Munir[1], F. Muzaffar[1], K. Shabbir[1], A. Imtiaz[1], H. Khan[1], E. Wohler[4], P. Witmer[4], N. L. Sobreira[5], G. Seo[6], R. J. Morell[7], T. B. Friedman[8]; [1]Univ Punjab, Lahore, Pakistan, [2]Bethesda, MD, [3]LMG, NIDCD, NIH, Bethesda, MD, [4]Johns Hopkins Univ., Baltimore, MD, [5]Johns Hopkins Univ, Baltimore, MD, [6]3billion, Seoul, Korea, Republic of, [7]NIDCD, NIH, Bethesda, MD, [8]NIDCD/NIH, Bethesda, MD

Disclosure Block: **S. Naz:** None.

Variants of at least 25 genes are associated with nonsyndromic recessively inherited moderate to severe hearing loss in human. We ascertained and studied 36 families with multiple individuals affected with either moderate hearing loss or hearing which progressively worsened to profound deafness. Families segregating variants of *GJB2* and the two non-coding variants of *HGF* (c.482+1986_482+1988delTGA, c.482+1991_482+2000del10) were identified using Sanger sequencing. Selected samples from affected and unaffected individuals in the remaining 33 families were exome sequenced. Altogether, we identified 12 novel and 25 reported variants in genes associated with either moderate-to-severe hearing loss or profound deafness. Variants of *SLC26A4* were associated with hearing loss in members of nine families. Variants of *CDH23* and *MYO15A* were each identified in three families. Additional likely pathogenic variants were observed in two families each (*ESPN*, *GJB2*, *MYO7A*, *OTOF*) or one family (*ELMOD3, EPS8*, *ESRRB*, *GIPC3*, *HGF*, *LHFPL5*, *MARVELD2*, *MY06*, *PCDH15*, *POU3F4*, *TECTA*, *TMPRSS3* and *TPRN*). Thirteen of the pathogenic alleles were missense variants, three were in-frame deletions and one was the reported three base pair non-coding deletion in *HGF*. Twenty of the identified variants affected splice sites or introduced frameshifts and stop codons in the reading frames of the respective genes. Thus, more than half of the alleles are predicted to affect gene expression or exon splicing. Currently, we are designing experiments to examine the pathogenic effects of some of these variants. Our research corroborates the contribution to moderate to severe hearing loss of variants of many previously studied â□□deafness genesâ□□ and for the first time implicates *LHFPL5*, *MY06* and *PCDH15* in the etiology of a moderate to severe hearing loss phenotype. This study was supported by the Baylor-Hopkins Center for Mendelian Genomics, 3billion inc., Higher Education Commission, Pakistan 3288 (SN) and funded (in part) by the National Institute on Deafness and Other Communication Disorders Intramural Research Program DC000086 to R.J.M., DC000039 to T.B.F.

PrgmNr 2228 - Inborn errors of immunity: Novel insights from four unrelated Indian families

View session detail

Author Block: **R. Lakshmi Priya**[1], P. H. Somashekar[1], D. L. Narayanan[1], A. Shrikiran[2], D. Rangaswami[3], A. Shukla[1]; [1]Med. Genetics, Kasturba Med. Coll., Manipal Academy of Higher Ed., Manipal, Udupi, India, [2]Dept. of Pediatrics, Kasturba Med. Coll., Manipal Academy of Higher Ed., Manipal, Udupi, India, [3]Dept. of Nephrology, Kasturba Med. Coll., Manipal Academy of Higher Ed., Manipal, Udupi, India

Disclosure Block: **R. Lakshmi Priya:** None.

Inborn errors of immunity are monogenic disorders of the immune system, which can present with recurrent infections, increased susceptibility to allergy, autoimmunity, autoinflammation, and malignancy. We present four Indian families with inborn errors of immunity, evaluated by exome sequencing. The proband in family 1, is a one-year-old male born to a second-degree consanguineous couple. He presented with failure to thrive, recurrent infections, and irritability. Laboratory investigations revealed macrocytic anemia, pancytopenia, and neutropenia. He had low IgG, IgA, IgM, CD3+ T cells and CD4+ T cells, indicating combined immunodeficiency. Singleton exome sequencing revealed a novel variant, c.1003C>T p.(Gln335Ter) in *TCN2* in the homozygous state, causing transcobalamin II deficiency (MIM#275350). The one-year-old proband in family 2 is the firstborn of second-degree consanguineous parents. He had recurrent infections and agammaglobulinemia. Singleton exome sequencing revealed a novel missense variant, c.374T>G p.(Val125Gln) in *CD79A* causing agammaglobulinemia 3 (MIM#613501). The thirteen-year-old female proband in family 3, was born to non-consanguineous parents and was evaluated for steroid-responsive membranous nephropathy. She also had short stature, hyperpigmentation of the skin, and a history of autoimmune hemolytic anemia and recurrent oral candidiasis. Her IgA and IgM levels were low. Mendeliome sequencing revealed a novel homozygous variant, c.520C>T p.(Arg124Ter) in *CARMIL2,* causing immunodeficiency 58 (MIM#618131). This is the first report of a subject with CARMIL2 deficiency presenting with membranous nephropathy. The proband in family 4, is a three-month-old child who presented with recurrent infections, sparse hair and eyebrows, hypopigmented skin, and small finger and toenails. Trio exome sequencing revealed a *de novo* heterozygous missense variant c.962A>G p.(His321Arg) in *FOXN1* causing T-cell lymphopenia, infantile, with or without nail dystrophy, autosomal dominant (MIM#618806). We herein add to the existing genotypic and phenotypic spectrum of the above-mentioned rare inborn errors of immunity. We also emphasize the role of next-generation sequencing in the rapid and efficient diagnosis of these disorders thus guiding treatment, genetic counseling, and prenatal diagnosis.

PrgmNr 2229 - Progressive familial intrahepatic cholestasis type3: two novel pathogenic variants in a cohort of Egyptian children with cholestatic disorders of infancy

View session detail

Author Block: **S. A. Sharaf**, N. N. Abbass, M. El-Sharkawy, M. S. El-Raziky, L. Hussein, H. Baz, M. H. Kamel, J. M. Medhat, H. El-Karaksy; Faculty of Med. Cairo Univ., cairo, Egypt
Disclosure Block: **S.A. Sharaf:** None.
Introduction: Progressive familial intrahepatic cholestasis (PFIC) is a heterogenous group of disorders in canalicular hepatobiliary transport results in progressive cholestasis and liver injury. Patients with PFIC type 3 have elevated gamma glutamyl transpeptidase and a variable degree of cholestasis that presents later in infancy or in early childhood. We aimed to identify mutations in the *ABCB4* gene responsible for PFIC3 in a group of infants and children presenting with high GGT chronic cholestasis. **Methodology:** This study was conducted at the Pediatric Hepatology Unit, Cairo University Hospital, Egypt over a 3 year period. Sixteen patients from both sexes presenting in the first year of life with cholestasis with high GGT were included. Patients were subjected to full history taking, detailed examination. Genetic analysis sanger sequencing of the coding exons and intronic boundaries of *ABCB4* gene was performed. Sequencing could be completed in nine patients only due to financial limitation. Allelic discrimination of *ABCB4:* p.(A1066D), rs31668 and rs31653 were done using Real-time PCR assay for all patients and 100 healthy controls. **Results** The age of disease onset ranged from birth till 12 months of age (median [IQR]: 0.6 (7) months). Positive consanguinity of the patients was 56.25%. Positive family history of similar conditions was 43.75%. The main presenting symptom was jaundice in 13 patients. GGT levels were elevated in all patients ranging from 112-792 U/L; 1 patient had mild elevation less than double the upper limit of normal and 10 patients were above five folds of the upper limit of normal. Nine patients had hypoalbuminemia and 5 had coagulopathy. Liver biopsy was performed in our entire study group. Their results varied from neonatal hepatitis, chronic hepatitis with variable degrees of fibrosis, bile duct proliferation suggestive of major bile duct obstruction and paucity of interlobular bile ducts. Two patients had novel pathogenic variant in homozygous pattern in exon 25 and a third patient in exon 27**.** One missense novel variant *ABCB4:* p.(A1066D) and the second novel variant is a single base substitution causes nonsense mutation and premature stop codon on the protein level which results in truncated protein in exon 27 *ABCB4*: p.(Q1181X) both classified as class 5 (pathogenic) according to the ACMG guidelines. **CONCLUSION:** The nonsense mutation is likely to present the symptoms of PFIC3 early in life. 2 Novel disease causing variants in exon 25 and 27 were detected in the Egyptian patients.

PrgmNr 2230 - Targeted exome sequencing identifies novel variants associated with non-syndromic hearing loss in an Indian family

View session detail

Author Block: **D. Kumari**, I. Panigrahi; Post Graduate Inst. of Med. Ed. and Res., Chandigarh, India
Disclosure Block: **D. Kumari:** None.
Introduction: Non -syndromic hearing loss (NSHL) is the most common sensory disorder worldwide and highly heterogeneous .Genetic factors account for approximately 50% congenital hearing loss cases. Multiple gene variants are associated with similar phenotype in case of non-syndromic hearing loss. Targeted exome sequencing has the advantage of studying multiple genes by massive parallel sequencing. Hence, it is advantageous for disorders in which multiple genes are involved in causation of similar clinical presentation.**Method**: The index case 10 years old, born of a non-consanguineous marriage with clinical indications of congenital profound sensorineural hearing loss. His ultrasound abdomen, thyroid profile and CMV serology were negative. His elder brother and maternal cousin are similarly affected and also undergone cochlear implant. Targeted next generation sequencing was performed covering for possible target genes of hearing loss using capture system as per the manufacturer's standard protocols. The libraries were sequenced to mean >80-100X coverage on Illumina sequencing platform. Sanger sequencing was performed to validate the identified variants in the index case and affected family members. Segregation analysis of the parents was done for identifying carrier state. **Result**: Molecular diagnosis confirmed by targeted exome sequencing revealed novel variants in the *TMC1* and *MYO7A* genes respectively. Two heterozygous variants NM_138691.3(*TMC1*):c.624C>A:p.Ser208Argand NM_(*MYO7A*):c.3856G>A:p.Ala1286Thr were identified in the brothers and homozygous variant NM_138691.3(*TMC1*):c.624C>A:p.Ser208Arg identified in the maternal cousin. On segregation analysis, parents of maternal cousin were confirmed carriers of the same variant. The *in silico* predictions of the variants were damaging by LRT and Mutation Taster 2.The identified variants were not reported in population databases 1000genome, gnomAD and ClinVar.**Conclusion**: Targeted NGS enabled the identification of novel variants in the multiply-affected family. Early diagnosis for congenital deafness helps in appropriate genetic counseling. High risk family members are better informed about the risk of their offspring being hearing impaired and can decide regarding further reproductive options. It also facilitates planning proper interventions for improving the hearing outcomes.

PrgmNr 2231 - A Novel Variant in RAP1GDS1 gene in a child with developmental

View session detail

Author Block: **H. Ahmed**[1], W. Eyaid[1], P. Bauer[2]; [1]KASCH Hosp., Riyadh, Saudi Arabia, [2]Centogene GmbH, Rostock, Germany
Disclosure Block: **H. Ahmed:** None.
ABSTRACT Background: 1-2% of children are affected with developmental delay all over the word. In more than half of them there is no clear cause. However genetic defect has been responsible for a quarter to one-half of identified cases. **Case presentation**: Here in our case we have found a 4yrs. Old child who presented with developmental delay, repetitive hand movements and dysmorphic features . Using whole exome sequencing we detected a homozygous variant in RAP1GDS1 gene in c.83del (p.LeuTrpfs*32), causing frame shift, likely affecting protein function (class 2P). Both parents are heterozygous carriers of the same variant. **Conclusion**: This gene has been shown to be expressed in several tissues including brain. Loss of function variants in the RAP1GDS1 are extremely rare in controls, and to our knowledge, not observed in homozygous state in any individual so we consider this c.83del (p.LeuTrpfs*32) variant in RAP1GDS1 gene as novel variant .

PrgmNr 2232 - Beta-Ketothiolase deficiency in a patient with coexistence of homozygous *ACAT1* gene mutation and homologous constitutional translocation t(4;11)(q13;q23) and father with XYY syndrome

View session detail

Author Block: **M. Albalwi**[1,2,3], E. AlOnazi[1], B. AlAwad[1], Y. AlThubati[1], A. AlKhaldi[1], W. Eyaid[1]; [1]King Abdulaziz Med. City, MNGHA, Riyadh, Saudi Arabia, [2]King Abdullah Intl., Med. Res. Ctr., King Saud bin Abdulaziz Univ. for Hlth.Sci., Riyadh, Saudi Arabia, [3]Coll. of Med., King Saud bin Abdulaziz Univ. for Hlth.Sci., Riyadh, Saudi Arabia

Disclosure Block: **M. Albalwi:** None.

Beta-ketothiolase deficiency is a rare metabolic disorder that is inherited in an autosomal recessive manner. Homologous constitutional chromosomal rearrangements is an extremely rarely reported chromosomal aberration. Here, we report on a 14 year-old boy whose parents are first degree consanguineous cousins. He has dysmorphic features, growth retardation, global developmental delay, bronchial asthma,and an Attention-Deficit/Hyperactivity disorder. Biochemical and molecular analysis *showed the presence of* homozygous c.410_418delinsT (p.Ser137Phrfs*37) *ACAT1* mutation. Both parents were heterozygous for the *ACAT1* gene mutation. *The chromosomal analysis showed* chromosomal homologous translocation t(4;11)(q13;q23) respectively. Besides that, the fatherâ□□s chromosomal analysis showed that he is a carrier for the translocation, also he has an XYY syndrome with 47,XYY,t(4;11)(q13;23). The motherâ□□s chromosomal analysis was 46,XX,t(4;11)(q13;23). Having a genetic disorder that is inconsistent with its usual clinical presentation should alert clinicians to the possibility of having another disorder, particularly in a highly consanguineous population. Proper clinical evaluation, and laboratory investigations are important for a proper genetic counseling.

PrgmNr 2233 - Biallelic *VPS35L* pathogenic variants cause 3C/Ritscher-Scinzel-like syndrome: Description of two novel cases confirming the pathogenicity and clinical diversity

View session detail

Author Block: **S. Otsuji**[1,2], K. Kato[1,2,3], C. HUBER LEQUESNE[4], S. Mizuno[2], M. RIO[4], S. Miyatake[5], Y. Nishio[1,3], N. Matsumoto[5], V. CORMIER-DAIRE[4], S. Saitoh[1]; [1]Dept. of Pediatrics and Neonatology, Nagoya City Univ. Graduate Sch. of Med. Sci., Nagoya, Japan, [2]Pediatrics and Clinical Genetics, Central Hosp., Aichi Dev.al Disability Ctr., Kasugai, Aichi, Japan, [3]Dept. of Pediatrics, Nagoya Univ. Graduate Sch. of Med., Nagoya, Japan, [4]DÃ©partement de GÃ©nÃ©tique MÃ©dicale (C.H.-L., V.C.-D.), Inst. Natl. de la SantÃ© et de la Recherche MÃ©dicale UnitÃ© 781, UniversitÃ© Paris Descartes-Sorbonne Paris CitÃ©, Inst. Imagine, HÃ´pital Necker Enfants Malades, Paris, France, [5]Dept. of Human Genetics, Yokohama City Univ., Graduate Sch. of Med., Yokohama, Japan

Disclosure Block: **S. Otsuji:** None.

Background VPS35L forms a retriever heterotrimer protein complex with VPS29 and DSCR3, which plays a pivotal role in the endosomal cargo recycling system, interacting with the WASH and CCC complex. Retriever complex promotes selective cell surface protein entry into the recycling pathway instead of degradation in lysosomes, and disruption of its function decreases expression level of numerous kinds of membrane proteins. We previously reported that *VPS35L* could be a novel cause for 3C-like syndrome, since a sibling with biallelic *VPS35L* loss-of-function variants showed overlapping symptoms of 3C syndrome. *CCDC22* and *WASHC5*, a member of CCC and WASHC complex respectively, were known to be the responsible genes for 3C syndrome. However, the association between VPS35L dysfunction and 3C-like syndrome remained unconfirmed because of the lack of other cases. **Methods** Exome sequencing was performed to identify pathogenic variants. Cellular biological analyses were carried out to prove the pathogenicity of the identified variants. **Results** We identified two novel cases with compound heterozygous loss-of-function variants in *VPS35L* (NM_020314.5), c.1650+2T>A; c.2982_2984del, p.Asn995del in the first case, and c.1577del; p.Ala526Valfs*14, c.3057del; p.Met1020Trpfs*2 in the other case. c.1650+2T>A and c.1557del frameshift variants were considered to induce nonsense mediated mRNA decay (NMD). In addition, although VPS35L-M1020-frameshift mutant was considered to escape NMD as it located C-terminal end, immunoprecipitation analysis of the VPS35L-M1020-frameshift mutant, as well as VPS35L-N995del, showed loss of binding ability to the endogenous VPS29, confirming that both cases have the biallelic loss of function variants in VPS35L. As with the case of previously reported sibling, both cases showed brachydactyly, skeletal complication, macrodontia, dysmorphic facial features, proteinuria, intellectual disability, and short stature. In addition, former case suffered from protein-losing-gastroenteropathy, and latter case showed hyperlipidemia, anterior ocular lesions and vascular malformations. **Conclusions** Biallelic *VPS35L* loss-of-function variants cause the 3C-like syndrome, with multisystem complications including what were not described in the conventional 3C syndrome.

PrgmNr 2234 - First reanalysis highlighting the increasing role of multiple molecular diagnoses in the field of rare diseases

View session detail

Author Block: **C. Racine**[1,2,3], A-S. DenommÃ©-Pichon[1,3], C. Engel[3], F. Tran-Mau Them[1,3], A-L. Bruel[3,1], A. Vitobello[1,3], A. Sorlin[1,3,2], H. Safraou[3], S. Nambot[1,2], J. Delanne[2], A. Garde[2,1], M. Bournez[2], S. Moutton[2,1], J. Thevenon[2], D. Lehalle[2], N. Houcinat[2], N. Jean Marcais[2], M. Willems[4], A. Verloes[5], F. Laffargue[6], L. Pinson[4], J. J. Lespinasse[7], E. Lacaze[8], D. Genevieve[4], O. PATAT[9], L. Lambert[10], M. Gerard Blanluet[11], C. Benigni[10], Orphanomix Physician's Group, V. Couturier[3], V. Bourgeois[3], P. Garret[3,1], C. Philippe[1,3], Y. Duffourd[1,3], L. Faivre[1,12], C. Thauvin-Robinet[1,13,3]; [1]Inserm UMR 1231 GAD, Genetics of Dev.al disorders, UniversitÃ© de Bourgogne-Franche ComtÃ©, FHU TRANSLAD, Dijon, France, [2]Ctr. de RÃ©fÃ©rence Maladies Rares 'Anomalies du DÃ©veloppement et syndromes malformatifs', FHU-TRANSLAD, Dijon, France, [3]UnitÃ© Fonctionnelle 'Innovation diagnostique dans les maladies rares' laboratoire de gÃ©nÃ©tique chromosomique et molÃ©culaire, Plateau Technique de Biologie, Dijon, France, [4]DÃ©partement de GÃ©nÃ©tique MÃ©dicale, Maladies rares et MÃ©decine PersonnalisÃ©e, CHU de Montpellier, Montpellier, France, [5]DÃ©partement de GÃ©nÃ©tique, Assistance Publique des HÃ´pitaux de Paris (AP-HP) HÃ´pital Robert DebrÃ©, Paris, France, [6]Service de GÃ©nÃ©tique mÃ©dicale, CHU de Clermont-Ferrand, Clermont-Ferrand, France, [7]Laboratoire de GÃ©nÃ©tique Chromosomique, CH GÃ©nÃ©ral, Chambery, France, [8]DÃ©partement de gÃ©nÃ©tique, CH du Havre, Le Havre, France, [9]Service de GÃ©nÃ©tique MÃ©dicale, HÃ´pital Purpan, CHU, Toulouse, France, [10]Service de GÃ©nÃ©tique Clinique, CHU de Nancy, Nancy, France, [11]Service de GÃ©nÃ©tique, CHU de Caen, Caen, France, [12]Ctr. de RÃ©fÃ©rence Maladies Rares 'DÃ©ficiences Intellectuelles de causes rares', FHU-TRANSLAD, Dijon, France, [13]Ctr. de RÃ©fÃ©rence Maladies Rares 'DÃ©ficiences Intellectuelles de causes rares', FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon, France
Disclosure Block: **C. Racine:** None.

Background: The etiological work-up of congenital abnormalities and intellectual disability (CA/ID) benefited from the implementation of next-generation sequencing technologies, which has opened the possibility of identifying multiple molecular diagnoses (MMD), ranging from 1.8% to 7.1% of positive cases in the literature. We aim to highlight the increasing role of MMD and their main characteristics and consequences in this field. Methods: We included patients referred to our diagnostic laboratory by local and Orphanomix reference centers for the etiological work-up of CA and/or ID, via exome sequencing (ES). The cohort was divided in two: a prospective cohort for ES data (re)analyzed from 01/2019 to 06/2021, and a retrospective cohort for local ES data (allowing easy access to clinical update & retro-phenotyping) analyzed prior to 2019; cut-off was based on preliminary encouraging results of prospective study of MMD from 01/2017 to 12/2020 combined with reanalysis of positive ES data rendered prior to 2017. Preliminary results: Out of the 604 positive ES data of the prospective cohort, 21 (3.5%) turned out being MMD and 20 (3.3%) included a variant of unknown significance (VUS) with a high probability of being reclassified (overall diagnostic rate of 6.8%). After reanalysis, 2 out of the 85 local positive ES data of the retrospective cohort, led to a MMD rate of 2.4%, and 5 VUS were deemed relevant, rising the overall diagnostic rate to 8.2%. Modes of transmission, types of variants, copy-number variants proportion, inheritance, contribution of VUS, temporality of MMD, and attributes of the patients will be discussed. Preliminary discussion: Rates of MMD, respectively 3.5% (6.8% including VUS) and 2.4% (8.2%) for the prospective and retrospective cohorts, are consistent with the literature and can be based on awareness of the interpretating team on this concept, up-to-date clinical data, knowledge enhancement on genes linked to human diseases, and bioinformatics tools improvement. The limits of our study are the potential missed MMD (by non-informative clinical updates, genes not yet known, VUS not yet reclassified, and ES limits). MMD appear to be underestimated and would benefit from greater insights since not only they have a

significant impact for patients and their families for genetic counselling, prenatal testing, and personalized follow-up, but also participate in better delineation of genesâ phenotypic spectra. This also raises the question whether positive ES data should benefit from reanalysis on request or systematically. Reanalysis of local positive ES data rendered in 2017/2018 and prospective study of 2021 first semester data are ongoing.

PrgmNr 2235 - Helsmoortel-van der Aa syndrome in a 14-year-old girl demonstrates intellectual impairment, facial dysmorphism and morbid obesity

View session detail

Author Block: **E. AL-Enezi**[1], W. Eyaid[2], K. Al-Anazi[3]; [1]Clinical Genetic and Metabolic Fellow, Dept. of Genetics, King Abdulaziz Med. City, Riyadh, Saudi Arabia, [2]Consultant Genetic and Endocrinology, Dept. of Genetics, King Abdulaziz Med. City, Riyadh, Saudi Arabia, [3]Pediatric Cardiology Fellow, Dept. of Pediatric Cardiology, King Abdulaziz Med. City, Riyadh, Saudi Arabia
Disclosure Block: **E. AL-Enezi:** None.
Helsmoortel-Van der Aa syndrome (HVDAS) (OMIM# 615873), is a recently discovered neurodevelopmental disorder, caused by pathogenic mutations in the ADNP gene (OMIM#611386), characterized by impaired intellectual development, motor delay, facial dysmorphism, autism spectrum disorder, visual difficulties, congenital heart diseases, gastrointestinal disorders, attention deficit/hyperactivity disorder (ADHD) and obsessive-compulsive behaviour. We report here a patient demonstrates the clinical features of Helsmoortel-van der Aa syndrome, molecularly confirmed, in addition to morbid obesity, short stature, bilateral ovarian cysts, solitary right kidney, distinctive dysmorphic features like highly arched eyebrows, proptosis and crowded teeth.

PrgmNr 2236 - Importance of genetic diagnosis on the PICU: case series during COVID-19 pandemic

View session detail

Author Block: **V. Plaiasu**, I. Apostol, A. Caia-Hoanas, A. Dobre, O. Farkas, A. Cochino, T. Ciomartan; INSMC Alessandrescu-Rusescu, Bucharest, Romania
Disclosure Block: **V. Plaiasu:** None.
Introduction: Congenital anomalies and genetic disorders are the most frequent cause of death in infancy. Children in intensive care frequently have a rare underlying genetic condition, often with atypical clinical presentation, unusual and puzzling symptoms, such as intractable seizures, acidosis or dysmorphic conditions with multiple congenital anomalies. It is critical to find the diagnosis of these children to reduce mortality and unessential intensive care and to decrease the anxiety of patient families. **Patients:** This study represents a retrospective evaluation for 30 unrelated children with age at recruiting ranged from 2 weeks to 3 years admitted to the PICU who were nominated between March 2020 and April 2021 during 14 months of COVID-19 pandemic. Data of the patients were obtained from their medical records from our paediatric tertiary care institution - PICU and genetics department. **Methods:** A molecular diagnosis was made by NGS technology: WGS (Whole Genome Sequencing) or WES (Whole Exome Sequencing) +/- CNV (Copy Number Variants) or targeted exome sequencing (TES) panels. Parental samples were not available to complete familial genetic investigation. **Results:** The diagnoses reported included pathogenic, likely pathogenic or uncertain (VUS) variants or incidental genetic findings of the genes such as ***TRAPPC4***, ***RYR1, MT-TS1, HEXB***, ***DOLK, ASL, UFM1, SLC4A1, ABCB11, TTC37, RAG1*** in a broad range of rare disorders including neurodegenerative conditions, myopathies, inherited metabolic disorders, congenital immunodeficiency and other syndromes. All positive results provided a valuable etiological explanation for the management of the cases and for the genetic counselling of the families. Changed specific clinical management of diagnosed patients was applied: avoiding morbidity from symptomatic monitor, avoiding severe infection, uncontrolled seizures, avoiding another accessory lab test, identifying unnecessary procedures for examination, furthering life-saving procedures. The molecular diagnosis of deceased infants was important due to implications for recurrence risk counselling. Also substantial challenges from medical and administrative point of view have been registered during pandemic period. **Conclusion**: The use of comprehensive genomic testing by NGS in intensively ill children facilitated diagnoses and assisted acute and long-term clinical decisions. Geneticists and critical care specialists should collaborate to get more accurate diagnoses for the patients with suspected monogenic disorders. Inborn errors of metabolism and neonatal encephalopathy represented the top diagnoses in intensive care setting.

PrgmNr 2237 - MLPA mediated molecular characterization of associated syndromes with orofacial cleft patients

View session detail

Author Block: **K. Avasthi**; Sanjay Gandhi Post Graduate Inst. of Med. Sci., Lucknow, India
Disclosure Block: **K. Avasthi:** None.
Introduction Orofacial clefts, also categorized as either Cleft Lip/Palate or Cleft Palate, are second most common congenital deformity. A worldwide birth prevalence orofacial clefts are 2.62 per 1000 live births. It can occur as part of mendelian syndromes or isolated/non-syndromic clefts. It is heterogeneous group of disorders, and Varies with ethnicity, gender and cleft type. It can arise through singleâ□□gene mutations, chromosomal abnormalities and effect of teratogens. The syndromic clefting designation refer to the presence of additional other physical and/or cognitive abnormalities, along with CL/P. **Materials and Methods** The present study was conducted on 25 syndromic orofacial cleft patients present with different clinical features. Genomic DNA was extracted from peripheral blood and multiplex ligation dependent probe amplification was performed for two probe sets, one for common microdeletions or second for Di-George syndrome. Sequence type electrophoresis was performed using ABI prism 310 genetic analyser. Later, fragment and comparative analysis part was carried out by coffalyser software. **Results** We detected 3 patients were with 22q11.2 deletion and shown that association of Di-George syndrome. No cases were detected for the other common microdeletion syndromes. **Conclusions** Our work demonstrated that Di-George Syndrome (22q11.2 deletion) is most common syndrome associated with orofacial cleft patients and it can be used as primary screening through the MLPA in orofacial cleft patients.

PrgmNr 2238 - Pulmonary valvular stenosis missed Rasopathies in two asymptomatic adults

View session detail

Author Block: **R. Louati**; FMS, Sfax, Tunisia
Disclosure Block: **R. Louati:** None.
Pulmonary stenosis (PS), rare among adults, develops usually before birth as a congenital heart defect (7-10%). PS ranges from a mild form to moderate severe forms valvular PS (VPS) is the most common type. Adults with PS may be asymptomatic regardless of the severity of their obstruction. Here, we report two adults diagnosed with a syndromic PS following the detection of a Noonan syndrome (NS) in their children. Material and methods: We analyzed by bidirectional sequencing all exons introns-exons boundaries of PTPN11 gene in two children suspected to have NS on the basis of facial dysmorphism, chest deformity congenital heart defect. They were affected respectively by an isolated VPS an associated VPS to an atrial septal defect. Results: The diagnosis of VPS associated to NS was confirmed. The first child had a c.923 A>G (p.N308S) mutation of PTPN11 exon 8 the second had a c.188 A>G (p.T63C) of PTPN11 exon 3. Familial investigation showed the presence of these mutations in the father of the first child the mother of the second. Clinical counselling showed a NS dysmorphism indicated cardiac echography which showed the presence of a silent VPS for the two 37-year-old parents. Conclusion: PS is often diagnosed in childhood, but sometimes it is not detected until later in life. This heart defect is commonly associated with congenital structural cardiac defect as tetralogy of Fallot Rasopathies like syndromes. It is primordial to look for to identify silent PVS missed Rasopathies in parents when a NS or other related Rasopathies are confirmed in their offsprings.

PrgmNr 2239 - The identification of microduplications in 17q21.31 locus that might disrupt the *KANSL1* gene

View session detail

Author Block: **A. Alaqeel**[1], A. Aladwani[2], A. Alahmad[2], A. Alhouly[2], H. AL-Balool[3]; [1]Kuwait Med. Genetic Ctr. (KMGC); Neurocare / 60 Degree Clinic, Sulaibikhat, Sabah Health Area, Kuwait, [2]Kuwait Med. Genetic Ctr. (KMGC), Sulaibikhat, Sabah Health Area, Kuwait, [3]Kuwait Med. Genetic Ctr. (KMGC), Sulaibikhat, Sabah Health Area, Kuwait
Disclosure Block: **A. Alaqeel:** None.
Background: 17q21.31 microdeletion syndrome, also known as Koolen-De Vries syndrome (KDVS), is a multisystem genetic disorder genetically recognized by the loss of function of the *KANSL1* gene and characterized by intellectual disability, hypotonia, failure to thrive, characteristic facial features, and other neurologic and cardiovascular defects. Point mutations in the *KANSL1* gene and microdeletions including the same gene cause KDVS.
Method: Copy number analysis using Affymetrix Cytoscan HD Microarray assay to analyse patients with unclear clinical diagnosis. It is to identify genetic aberrations in these cases. Results: The oligonucleotide array analysis detected a gain in 17q21.31 region (approximately chr17:44187491-44784639) in 5 unrelated patients; which are approximately 570-600 kb in size and includes 8 genes. Although this type of aberration has not been reported as causative for a any known microduplication syndrome; it partially overlaps with the *KANSL1* gene and may potentially compromise its function. Due to high sensitivity of Affymetrix Cytoscan HD Array platform; it is possible to detect the duplication that might cause disruption in the structure of *KANSL1* gene. As a result, it may lead to loss-of-function of *KANSL1* gene. Moreno-Igoa *et al* (2015) reported a patient with balance *de novo* translocation between chromosome 1 and 17 (46, XX,t(1;17)(q12;q21) that includes duplicated segment that disrupts the *KANSL1* gene. Two out of the five cases also harbor a pathogenic aberration, the first one with 22q11.21 deletion (DiGeorge syndrome), and the second with 7q11.23 deletion (Williams Syndrome). Although both cases do exhibit clinical features related to their syndromes, other features however were detected that overlap with 17q21.31 microdeletion syndrome. Microduplication in 17p21.31 region that includes the first three exons of *KANSL1* gene reported to be with associated with 22q11.2 deletion and was found to significantly influence the patients clinical phenotype (Leon *et al* 2016). Conclusion: The investigation of the human genome of undiagnosed clinical genomic cases using microarray has identified microduplication located in 17q21.31 that overlap partially with the *KANSL1* gene in 5 unrelated individuals. Although there is some evidence supporting the pathogenicity of a microduplication disrupting the *KANSL1* gene function; however, more evidence is needed to re-classify this aberration. Therefore, further testing is required to investigate this aberration. Thus, our laboratory classifies this type of aberration as variant of unknown significance according to *American College of Medical Genetics* guidelines.

PrgmNr 2240 - Clinical whole genome sequencing as an efficient first-tier diagnostic test for patients withrare diseasesfrom a resource-limited setting in the Democratic Republic of Congo

View session detail

Author Block: **A. Lumaka Zola**[1,2], G. Mubungu[1], A. Malhotra[3], S. Sajan[3], P. Lukusa[1], P. Makay[1], M. Ngole[1], D. Tshika[1], .. ILS Interpretation, Reporting, and Bioinformatics[3], D. L. Perry[3], J. W. Belmont[3], J. Ortega[3], R. J. Taft[4]; [1]UniversitÃ© de KINSHASA, KINSHASA, Congo, Democratic Republic of the, [2]UniversitÃ©de LiÃ¨ge, LiÃ¨ge, Belgium, [3]Illumina Inc., San Diego, CA, [4]Illumina Inc, San Diego, CA
Disclosure Block: **A. Lumaka Zola:** None.
There is increasing utilization of clinical whole genome sequencing (cWGS) as a first-tier diagnostic test for rare and undiagnosed diseases in Western countries. Thus far, this approach has not been evaluated in Africa. The Illumina iHope Program is a philanthropic initiative that aims to provide cWGS to individuals with a suspected rare genetic disease who have limited access to molecular testing. Recently, this program expanded to provide cWGS to a resource-limited and underserved population from the Democratic Republic of Congo through an ongoing partnership with the University of Kinshasa. The initial cohort includes 40 patients ranging in age from 10 months to 78 years from 35 unrelated families. Following cWGS, variants were reported in 24 of the 35 families (68.6%). Fifteen of the 35 families (42.9%) carried at least one likely pathogenic or pathogenic variant identified in a gene associated with a disease that was consistent with the phenotype. An additional nine individuals (25.7%) carried variants of uncertain significance due to currently limited evidence, but showed some clinical congruence to the reported phenotype. Clinically significant variants included single nucleotide variants (SNV) and small indels (n=12), aneuploidy (n=1), copy number variants (CNV; n=4), and one repeat expansion. Five individuals had multiple variant types reported, most often a compound heterozygous pair with a CNV and SNV. Of note, two individuals with thrombocytopenia-absent radius syndrome had an ~448 kb deletion of 1q21.1 in trans with a recently described hypomorphic variant in the 3â□□ UTR of *RBM8A* that is common in the African population. In addition, variants were identified that were informative to multiple affected family members including three siblings with suspected retinitis pigmentosa who all carried biallelic variants in *ADAM9* (associated with cone rod dystrophy). Similarly, three siblings with spinocerebellar ataxia all had a pathogenic repeat expansion in the *ATXN7* gene. The cWGS results also provided information on potential change in clinical management in two individuals: monitoring for gastric cancer in an individual with a *CDH1* variant and possible treatment with levothyroxine in an individual with a *THRA* variant, based on literature evidence showing beneficial outcomes of treatment in individuals with the variant. These data demonstrate the comprehensive diagnostic advantage of cWGS as a first-tier test in individuals with a suspected genetic disease in resource-limited African populations. Gene and variant information from this unique and under-represented population will be shared with the community through ClinGen and ClinVar.

PrgmNr 2241 - Current status of genetic diagnosis laboratories and frequency of genetic variants associated with cystic fibrosis through a newborn-screening program in Turkey

View session detail

Author Block: **S. Tug Bozdogan**[1,2], C. Mujde[2], I. Boga[1,2], O. Sonmezler[2], A. Hanta[2], C. Rencuzogullari[2], D. Ozcan[3], D. Ufuk Altintas[3], A. Bisgin[1,2]; [1]Cukurova Univ., Faculty of Med., Dept. of Med. Genetics, Adana, Turkey, [2]Cukurova Univ., AGENTEM (Adana Genetic Diseases Diagnosis and Treatment Ctr.), Adana, Turkey, [3]Cukurova Univ., Faculty of Med., Dept. of Pediatric Allergy and Immunology, Adana, Turkey
Disclosure Block: **S. Tug Bozdogan:** None.
Background: Cystic fibrosis (CF) is the most common worldwide, life-shortening multisystem hereditary disease, with an autosomal recessive inheritance pattern caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. The national newborn screening (NBS) program for CF has been initiated in Turkey since 2015. If the immunoreactive trypsinogen (IRT) is elevated (higher than 70 Î¼g/L in the second control) and confirmed by sweat test or clinical findings, genetic testing is performed. The aims of this study are to emphasize the effect of NBS on the status of genetic diagnosis centers with the increasing numbers of molecular testing methods, and to determine the numbers and types of CFTR mutations in Turkey. **Methods:** The next-generation sequencing (NGS) and multiplex ligation-dependent probe amplification (MLPA) results of 1595 newborns, who were referred to Cukurova University Adana Genetic Diseases Diagnosis and Treatment Center (AGENTEM) for molecular genetic testing, were evaluated with positive CF NBS program results since 2017. **Results:** According to the results; 560 (35.1%) of the 1595 patients carried at least 1 (one) CF-related variant, while 1035 patients (64.9%) had no mutation. Compound heterozygosity for two mutations was the most common in patients, while two detected variants were homozygote in 14 patients. A total of 161 variants were detected in 561 patients with mutations. Fifteen novel variants that have not been previously reported were found. Moreover, p.L997F was identified as the most frequent pathogenic mutation that might affect the IRT measurements used for the NBS. The distribution of mutation frequencies in our study showed a difference from those previously reported; for example, the well-known p.F508del was the third most common (n = 42 alleles), rather than the first. The most striking finding is that 313 cases had a pathogenic variant together with the V470M variant, which might have a cumulative effect on CF perpetuation. **Conclusion:** This study is the first to determine the mutational spectrum of CFTR in correlation with the NBS program in the Turkish population. NBS for CF raises issues regarding screening in diverse populations, both medical and non-medical benefits, and carrier identification. Through the lens of NBS, we focused on the integrated diagnostic algorithms and their effect on the results of genetic testing.

PrgmNr 2242 - Detection of pathogenic CNVs in isolated and non-isolated congenital heart defects (CHD) by MLPA and CMA (retrospective cohort study)

View session detail

Author Block: **P. Capkova**[1,2], Z. Capkova[1,2], J. Srovnal[3], S. Travnickova[2], V. Curtisova[1,2], A. Stefekova[1], J. Petrkova[1,2], E. Klaskova[4,2]; [1]Dept. of Med. Genetics, Univ. Hosp. Olomouc, Olomouc, Czech Republic, [2]Faculty of Med. and Dentistry, Palacky Univ. Olomouc, Olomouc, Czech Republic, [3]Inst. of Molecular and Translational Med., Palacky Univ. Olomouc, Olomouc, Czech Republic, [4]Dept. of Paediatrics, Univ. Hosp. Olomouc, Olomouc, Czech Republic
Disclosure Block: **P. Capkova:** None.
CHDs are among the commonest birth defects. Pathogenic CNVs are frequent causes of CHD. The aim of the study was to determine the detection rate of pathogenic CNVs in prenatal and postnatal cohorts of isolated and non-isolated CHD cases after exclusion of aneuploidies by MLPA and CMA. **Method:** 129 samples (peripheral blood, amniotic fluid, chorionic villi) of patients with CHD were tested by MLPA and/or aCGH. **Results**: We detected 6 (13.33 %) and 12 (14.29 %) pathogenic CNVs in prenatal (45) and postnatal (84) cohort respectively. The most frequent were CNVs of 22q11.2 locus: they were present in one third of the prenatal as well postnatal samples and segregated with various types of CHD, including minor abnormalities (FOA etc.). Only one pathogenic CNV (1.92 %) and 5 VOUS CNVs (9.62 %) were detected in the cohort with isolated CHD (52) compared to 17 (22.08 %) pathogenic and 3 (3.9 %) VOUS CNVs in the group where CHD was present together with other congenital malformations or comorbidities (77). In addition to the recurrent microdeletions which are a known cause of CHD (1p36.33, 1q21.1, 4p16.3p16.2, 5p15.33, 7q11.2, 8p23.1, 9p34.3) we also detected two large regions of homozygosity - one resulting from upd(20)mat and one on chromosome 10 as a consequence of parental consanguinity. This finding suggests a possible role of AR condition or imprinting in CHD. Incidentally we also detected a point mutation in *BMP4 (*rs897543876) in a family with bicuspid aortic valve and aortic dilatation by MLPA testing and Sanger sequencing. **Conclusion**: Targeted (MLPA) and whole-genome (CMA) investigation of pathogenic CNVs is beneficial especially where other co-morbidities/malformations accompany CHD. The rate of detection of pathogenic CNVs is modest in isolated CHD cases where polygenic aetiology is assumed. High frequency of microdeletion 22q11.2 in both cohorts and its highly variable phenotype (wide spectrum of CHD, some of them undetectable by prenatal US examination, absence of CHD) suggests that it might be helpful to perform screening for the microdeletion routinely in prenatal diagnosis. Supported by MH CZ-DRO(FNOl,00098892), IGA_LF_2021_019.

PrgmNr 2243 - Genetic determinants of mosaic loss of the X chromosome in peripheral leukocytes of 395,036 women from 3 biobanks

View session detail

Author Block: **A. Liu**[1], G. Genovese[2,3], Y. Zhao[4], FinnGen, P-R. Loh[5,3], A. Ganna[1,3,6], J. R. Perry[4], M. J. Machiela[7]; [1]Inst. for Molecular Med. Finland (FIMM), Univ. of Helsinki, Helsinki, Finland, [2]Stanley Ctr., Broad Inst. of Harvard and MIT, Cambridge, MA, [3]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, [4]MRC Epidemiology Unit, Inst. of Metabolic Sci., Univ. of Cambridge, Cambridge, United Kingdom, [5]Brigham and Women's Hosp. / Harvard Med. Sch., Boston, MA, [6]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [7]Natl. Cancer Inst., Rockville, MD

Disclosure Block: **A. Liu:** None.

Mosaic loss of the X chromosome (mLOX) is the most frequently occurring age-related somatic chromosomal alteration detected in peripheral leukocyte DNA of females, with a prevalence of 3.9% in women younger than 50 and reaching 33.3% after age 70. However, little is known about the genetic causes of mLOX and its epidemiological consequences. Building on success from prior genome-wide association studies (GWAS) of mosaic Y loss (mLOY) in males, we performed a GWAS meta-analysis of mLOX in 395,036 women of European ancestry (36,178 with mLOX) from UK Biobank, FinnGen, and Mass General Brigham Biobank to characterize germline genetic architecture of mLOX.

We identified 8 autosomal germline susceptibility loci, including one locus in the MHC region and multiple loci located near genes associated with blood cell counts (*SP140L*, *SCML4*, *MSC-AS1*, *CTD-2207O23.3* and *KRI1*), blood protein level (*ADAMTS5*), and chronic lymphocytic leukemia (*SP110*). Rare variant analyses in UK Biobank exome sequencing data (N=101,027) identified additional rare variation in the *BUB1B* cancer predisposition gene (a mitotic checkpoint kinase) associated with increased risk of mLOX ($P=5.4x10^{-7}$). Only one of the genome-wide significant signals for mLOX (MHC locus) was associated with mLOY ($P=1.7x10^{-6}$) suggesting distinct molecular drivers for mLOX relative to mLOY. However, the genetic correlation between mLOX and mLOY was 0.25 (95% CI 0.08 to 0.42), indicating some shared genetic architecture from common variants.

Allelic shift analyses were performed on X chromosome data to identify germline variants for which one allele was preferentially lost in mLOX cases with heterozygous genotypes. Multiple such variants were identified over a large region spanning the centromere (P-250) as well as near chromosome X genes associated with skewed X-inactivation (*PLS3* and *ITM2A*) and cancer risk/progression (*FAM9C*, *CT45A1* and *SAGE1*). Furthermore, many of these allelic shift variants were near genes that modulate blood cell counts (*P2RY8*, *WAS*, *PJA1*, *PLS3*, *ITM2A* and *TMEM255A*). As mLOX preferentially involves the inactive X chromosome, such associations hint towards some form of competitive advantage for clonal growth or selection among the two populations of cells inactivating different X chromosomes. Leveraging genotype data from 400k females, we detected multiple germline variants associated with mLOX suggesting susceptibility to mLOX exhibits relationships with skewed X inactivation, blood cell traits, and cancer predisposition genes. Allelic shift analyses further demonstrate the strong cis selection of specific X variants providing novel insights into the genetic etiology of mLOX.

PrgmNr 2244 - Low-pass genome sequencing-based detection of absence of heterozygosity: validation in clinical cytogenetics

View session detail

Author Block: **Z. Dong**[1], M. Chau[1], Y. Zhang[1], Z. Yang[1], m. Shi[2], Y. Wah[1], Y. Kwok[1], T. Leung[3], C. C. Morton[4], K. choy[5]; [1]The Chinese Univ. of Hong Kong, Hong Kong, Hong Kong, [2]Hong Kong, China, [3]The Chinese Univ. of Hong Kong, Hong Kong, China, [4]Brigham & Women's Hosp, Boston, MA, [5]
Disclosure Block: **Z. Dong:** None.
Introduction: Absence of heterozygosity (AOH) is a genetic characteristic known to cause human genetic disorders through autosomal recessive or imprinting mechanisms. However, the analysis of AOH via low-pass genome sequencing (GS) is not yet clinically available. **Materials and Methods:** Low-pass GS (fourfold) with different types of GS libraries was performed on 17 clinical samples with previously ascertained AOH by chromosomal microarray analysis (CMA). In addition, AOH detection was performed with low-pass GS data in 1,639 cases that had both GS and high-probe density CMA data available from the 1000 Genomes Project. Cases with multiple AOHs (coefficient of inbreeding F â¥ 1/32) or terminal AOHs â¥5 Mb (suspected uniparental disomy [UPD]) were reported based on the guidelines of the American College of Medical Genetics and Genomics. **Results: We first demonstrated the optimal read-depth for AOH analysis to be fourfold regardless of sequencing modes (paired-end or single-end) and library types (small-insert or large-insert). In addition,** low-pass GS revealed suspected segmental UPD and multiple AOHs (F â¥ 1/32) in nine and eight clinical cases, respectively, consistent with CMA. Lastly, among the 1,639 samples with CMA and GS available in the 1000 Genomes Project, low-pass GS not only consistently detected multiple AOHs (F â¥ 1/32) in 18 cases, but also reported 60 terminal AOHs in 44 cases including four mosaic AOHs at a level ranging from 50% to 75%. **Conclusion:** Overall, our study demonstrates the feasibility of AOH analysis (â¥5 Mb) with low-pass GS data and shows high concordance compared with CMA.

PrgmNr 2245 - Multilocus disease-causing genomic variations for Mendelian disorders: role of systematic phenotyping and implications on genetic counselling

View session detail

Author Block: **D. Narayanan**[1], D. Udyawar[1], P. Kaur[1], S. Sharma[2], N. Suresh[2], S. Nampoothiri[3], M. do Rosario[1], P. Somasekhar[1], L. Rao[1], N. Kausthubham[1], P. Majethia[1], S. Pande[1], R. Bhat Y[4], S. Aroor[4], S. L. Bielas[5], G. Katta[1], A. Shukla[1]; [1]Kasturba Med. Coll., Manipal, MAHE, Manipal, Manipal, India, [2]Lady Hardinge Med. Coll. and Kalawati Saran Childrenâ□□s Hosp., New Delhi, India, [3]Amrita Inst. of Med. Sci. and Res. Ctr., Cochin, India, [4]Dept. of Pediatrics, Kasturba Med. Coll., Manipal, MAHE, Manipal, Manipal, India, [5]Univ. of Michigan, Ann Arbor, MI

Disclosure Block: **D. Narayanan:** None.

Multilocus disease-causing genomic variations (MGVs) and multiple genetic diagnoses (MGDs) are increasingly being recognised in individuals and families with Mendelian disorders. This can be mainly attributed to widespread use of genomic tests for evaluation of these disorders. We conducted a retrospective study of families evaluated over the last 6 years at our centre to identify families with MGVs and MGDs. MGVs were observed in fourteen families. We observed five different consequences: (i) Individuals with MGVs presenting as blended phenotypes (ii) Individuals with MGVs presenting with distinct phenotypes (iii) Individuals with MGVs with age-dependent penetrance (iv) Individuals with MGVs with one phenotype obscured by another more predominant phenotype (v) Two distinct phenotypes in different individuals in families with MGVs. Consanguinity was present in eight (8/14, 57.1%) of them. Thirteen families had two Mendelian disorders and one had three Mendelian disorders. The risk of recurrence of one or more conditions in these families ranged from 25% to 75%. Our findings underline the importance of the role of a clinical geneticist in systematic phenotyping, challenges in genetic counselling and risk estimation in families with MGVs and MGDs, especially in highly inbred populations.

PrgmNr 2246 - The high-risk phenotypes of genetic disease in a Neonatal Intensive Care Unit population from the China neonatal genomes project

View session detail

Author Block: **T. Xiao**[1], N. Qi[1], H. Chen[1], H. Wang[2], L. Yang[2], B. Wu[1], Y. Cao[1], G. Cheng[1], L. Wang[1], H. Mei[1], Y. Lu[3], X. Dong[4], W. Zhou[2]; [1]Childrenâ□□s Hosp. of Fudan Univ., Shanghai, China, [2]Children's Hosp. of Fudan Univ., Shanghai, China, [3]Children Hosp. of Fudan Univ., shanghai, China, [4]Fudan Univ., Shanghai, China

Disclosure Block: **T. Xiao:** None.

It is difficult to determine when neonates required genetic testing, especially for the neonates without special facial features or multiple congenital anomalies. Therefore, we conducted an observational study to systematically analyze the phenotypes of the genetic disease in a Neonatal Intensive Care Unit population between June 1, 2016 and June 30, 2020. We systematically investigated the phenotypes of the neonates with the Auto-Neo-HPO pipeline assistance and compared the phenotypes between the neonates with or without positive genetic diagnosis. Of 2,600 enrolled neonates, 248 neonates (9.5%) were diagnosed with a positive genetic diagnosis. The most common phenotypes in neonates with positive genetic diagnosis was abnormal heart morphology (HP: 0001627) (50.8%, 126/248), followed by sepsis (HP:0100806) (34.2%, 85/248), jaundice (HP:0000952) (31.8%, 79/248), encephalopathy (HP:0001298) (24.6%, 61/248), seizure (HP:0001250) (23.8%, 59/248). Compared to the neonates with negative genetic diagnosis, the risk phenotypes included muscular hypotonia (HP:0001252) (p

PrgmNr 2247 - A novel splice site variant causes SLC13A5-related developmental and epileptic encephalopathy in large consanguineous family

View session detail

Author Block: **H. Aldhalaan**[1], H. AlQudairy[1], M. AlNakiyah[2], N. Almutairi[2], R. Alghofaili[2], S. Alruways[2], A. Albakheet[3], M. Aldosary[4], D. Colak[5], N. Kaya[5]; [1]KFSHRC, Riyadh, Saudi Arabia, [2]KSU, Riyadh, Saudi Arabia, [3]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, [4]KFSH&RC, Riyadh, Saudi Arabia, [5]KFSH & RC, Riyadh, Saudi Arabia
Disclosure Block: **H. Aldhalaan:** None.
SLC13A5 (solute carrier family 13, member 5) encodes Na+ coupled citrate (NaCT) transporter protein which can carry dicarboxylates in addition to tricarboxylates substrates. Mutant *SLC13A5* proteins can lose the ability to bind to Na molecules, therefore failing to transport citrate, an essential compound for brain function, to from extracellular matrix to cytosol. Lack of cellular citrate results in energy deficit in the brain, thus contributes to pathogenesis of epilepsy and delayed brain development. Biallelic mutations in *SLC13A5* (homozygous or compound heterozygous) are associated with different phenotypic features such as seizures, intellectual disability, global developmental delay, and tooth dysplasia. Up to date 36 different disease-causing variants have been reported in the literature. Here, we present a novel splice site variant in three Saudi patients from a large consanguineous family. Patients experienced epileptic encephalopathy, microcephaly, motor difficulties, and intellectual disability. Most treatment regimen of such cases consists of group of traditional antiepileptic drugs (e.g. diazepam, topiramate, and depakine) to control their seizures. *SLC13A5* related epilepsy can be treated with Stiripentol (second-generation antiepileptic drug) not only to control seizures but also to enhance cellular citrate levels. Our study expands genetic spectrum of the disease and presents alternative approach to the suffering family to seek healthy siblings though pre-implementation genetic diagnosis.

PrgmNr 2248 - Apolipoprotein E4 and meningeal lymphatics in Alzheimer disease: a conceptual framework

View session detail

Author Block: **A-F. Mentis**, G. Chrousos; Univ. Res. Inst. of Child and Maternal Hlth.& Precision Med., Athens, Greece
Disclosure Block: **A. Mentis:** None.
The potential existence and roles of the meningeal lymphatic system in normal and pathological brain function have been a long-standing enigma. Recent evidence suggests that meningeal lymphatic vessels are present in both the mouse and human brain; in mice, they seem to play a role in clearing toxic amyloid-beta peptides, which have been connected with Alzheimer disease (AD). Here, we review the evidence linking the meningeal lymphatic system with human AD. Novel findings suggest that the recently described meningeal lymphatic vessels could be linked to, and possibly drain, the efferent paravascular glial lymphatic (glymphatic) system carrying cerebrospinal fluid, after solute and immune cell exchange with brain interstitial fluid. In so doing, the glymphatic system could contribute to the export of toxic solutes and immune cells from the brain (an exported fluid we wish to describe as glymph, similarly to lymph) to the meningeal lymphatic system; the latter, by being connected with downstream anatomic regions, carries the glymph to the conventional cervical lymphatic vessels and nodes. Thus, abnormal function in the meningeal lymphatic system could, in theory, lead to the accumulation, in the brain, of amyloid-beta, cellular debris, and inflammatory mediators, as well as immune cells, resulting in damage of the brain parenchyma and, in turn, cognitive and other neurologic dysfunctions. In addition, we provide novel insights into APOE4-the leading genetic risk factor for AD-and its relation to the meningeal lymphatic system. In this regard, we have reanalyzed previously published RNA-Seq data to show that induced pluripotent stem cells (iPSCs) carrying the APOE4 allele (either as APOE4 knock-in or stemming from APOE4 patients) express lower levels of (a) genes associated with lymphatic markers, and (b) genes for which well-characterized missense mutations have been linked to peripheral lymphedema. Taking into account this evidence, we propose a new conceptual framework, according to which APOE4 could play a novel role in the premature shrinkage of meningeal lymphatic vessels (meningeal lymphosclerosis), leading to abnormal meningeal lymphatic functions (meningeal lymphedema), and, in turn, reduction in the clearance of amyloid-beta and other macromolecules and inflammatory mediators, as well as immune cells, from the brain, exacerbation of AD manifestations, and progression of the disease. Altogether, these findings and their potential interpretations may herald novel diagnostic tools and therapeutic approaches in patients with AD.

PrgmNr 2249 - ÂÂÂÂASL deficiency in ALDH1A1^{+}- neurons in the substantia nigra metabolically promotes neurodegenerative phenotypes

View session detail

Author Block: **A. Erez**[1], S. Lerner[2], S. Nagamani[3], J. Baruteau[4]; [1]Weizmann Institue of Sci., Rehovot, Israel, [2]Weizmann, Rehovot, Israel, [3]Baylor Coll. of medicine, Houston, TX, [4]Great Ormond Street Inst. of Child Hlth., London, United Kingdom

Disclosure Block: **A. Erez:** None.

Argininosuccinate lyase (ASL) is essential for the NO-dependent regulation of tyrosine hydroxylase (TH) and thus for catecholamine production. Using a conditional mouse model with loss of ASL in catecholamine neurons, we demonstrate that ASL is uniquely expressed in ALDH1A1^{+} subpopulation of dopaminergic neurons in the substantia nigra pars compacta, that are pivotal for the pathogenesis of Parkinson disease (PD). Neuronal loss of ASL results in catecholamine deficiency, in accumulation and formation of tyrosine aggregates, in elevation of Î±-synuclein, and phenotypically in motor and cognitive deficits. NO supplementation rescues the formation of aggregates as well as motor deficiencies. Our data point to a potential metabolic link between accumulations of tyrosine and the seeding of pathological aggregates in neurons as initiators for the pathological processes involved in neurodegeneration. Hence, interventions in tyrosine metabolism *via* regulation of NO levels may be therapeutic benefits for the treatment of catecholamine-related neurodegenerative disorders.

PrgmNr 2250 - Compound heterozygous *ATM* variants cause late onset cerebellar and extrapyramidal disease in the absence of telangiectasia in a consanguineous Pakistani family

View session detail

Author Block: **F. Aslam**[1,2], W. Ji[2], L. Jaffries[2], A. Wadhwa[3], M. Konstantino[2], M. Noman[1], T. Ishaq[1], N. S. Bamford[3,4], S. Naz[1], S. A. Lakhani[2]; [1]Sch. of Biological Sci., Univ. of the Punjab, Lahore, Pakistan, Lahore, Pakistan, [2]Pediatric Genomic Discovery Program, Dept. of Pediatrics, Yale Univ. Sch. of Med., New Haven, CT, USA, New Haven, CT, [3]Dept. of Neurology, Yale Univ. Sch. of Med., New Haven, CT, USA, New Haven, CT, [4]Dept.s of Pediatrics and Molecular and Cellular Physiology, Yale Univ. Sch. of Med., New Haven, CT, USA, New Haven, CT

Disclosure Block: **F. Aslam:** None.

Ataxia-telangiectasia (A-T) is a heterogeneous genetic disorder with a recessive mode of inheritance resulting from homozygous or compound heterozygous variants in the A-T mutated gene (*ATM)*. ATM is a serine-threonine kinase that is activated by DNA double strand breaks and then initiates a number of signaling cascades modulating a variety of processes including cell cycle, DNA repair, and cell death. Dysfunctional ATM leads to A-T, which is marked by compromised immunity, chromosomal instability and increased risk of malignancies. We present a consanguineous Pakistani family with multiple individuals having late-onset ataxia. Affected individuals manifested gait and limb ataxia, postural instability and generalized dystonia. Whole exome sequencing identified a known pathogenic nonsense *ATM* variant c.2413C>T, p.(Arg805Ter) in trans with a previously unreported missense variant c.8708C>T p.(Pro2903Leu). Sanger sequencing of DNA from all participants revealed that the variants segregated with the disease in this family. A few cardinal features of A-T were absent that include telangiectasia of the eyes and skin, which led to the initial misdiagnosis of the disease as cerebellar ataxia. There were no reports of malignancies in the family and affected individuals were alive in their third and fourth decades of life, which is uncommon for A-T. Thus, molecular analysis of affected individuals resulted in reclassification of the disease as A-T, an example of reverse phenotyping. We suspect that this novel missense variant likely has residual kinase activity, thus causing a milder form of late-onset A-T with the absence of telangiectasia. Our findings extend the allelic spectrum of *ATM* variants and expand phenotypic heterogeneity of A-T. This study was funded by Higher Education Commission (HEC) of Pakistan with the award of IRSIP scholarship to F.A. and research grant HEC 2877 to S.N. and by a Yale New Haven Hospital research grant to S.A.L.

PrgmNr 2251 - Dysregulated expression levels of *APH1B* in peripheral blood are associated with brain atrophy and amyloid-Î² deposition in Alzheimerâ□□s disease

View session detail

Author Block: **Y. Park**[1], J-M. Pyun[1], A. Hodges[2], J-W. Jang[3], P. J. Bice[4], S. Kim[1], A. J. Saykin[4], K. T. Nho[4]; [1]Seoul Natl. Univ. Bundang Hosp., Seongnam-si, Korea, Republic of, [2]Kingâ□□s Coll. London, London, United Kingdom, [3]Kangwon Natl. Univ. Hosp., Chuncheon-si, Korea, Republic of, [4]Indiana Univ. Sch. of Med., Indianapolis, IN

Disclosure Block: **Y. Park:** None.

The interaction between brain and periphery might play a crucial role in the development of Alzheimer's disease (AD). Using blood transcriptomic profile data from two independent AD cohorts, we performed expression quantitative trait locus (*cis*-eQTL) analysis of 29 significant genetic loci from a recent large-scale genome-wide association study to investigate the effects of the AD genetic variants on gene expression levels and identify their potential target genes. We then performed differential gene expression analysis of identified AD target genes and linear regression analysis to evaluate association of differentially expressed genes with neuroimaging biomarkers. *cis*-eQTL analysis identified and replicated significant associations in seven genes. *APH1B* expression levels in blood increased in AD and were associated with entorhinal cortical thickness and global cortical amyloid-Î² deposition. An integrative analysis of genetics, blood-based transcriptomic profiles, and imaging biomarkers suggests that *APH1B* expression levels in blood might play a role in the pathogenesis of AD.

PrgmNr 2252 - IMMP2L gene and Gilles de la Tourette syndrome a new case & review of the literature

View session detail

Author Block: **N. Almobadel**, H. Elghezal, I. Ben Abdallah, a. el onazi, A. Alhashem, O. Al Suhaibani; Prince Sultan military Med. city, Riyadh, Saudi Arabia
Disclosure Block: **N. Almobadel:** None.
Gilles de la Touretteâ□□s syndrome (GTS) is an inherited neuropsychiatric disorder characterized by the presence of chronic multiple motor and vocal tics affecting up to 1% of schoolchildren with a wide range of severity. The etiology of GTS is complex, combined with multiple genes and environmental factors. Family and twin studies have conï¬□rmed that the majority of GTS cases are inherited, although the exact mode of inheritance is not yet known. In recent years, several new candidate genes, in particular SLITRK1, IMMP2L, CNTNAP2, and NLGN4 genes, were identiï¬□ed through linkage studies and structural genomic aberrations, as major candidate genes in GTS. 7q31 Chromosomal aberrations studies have identiï¬□ed rare structural aberrations associated with GTS in particular in a case were the Inner mitochondrial membrane protein 2L (IMMP2L) gene was disrupted by a deletion of exons 1-3 in a male patient with GTS. We report here a 3 years old male, of a first degree cousins unaffected parents with no family history, presenting a neonatal form of GTS with recurrent abnormal movements diagnosed since the age of 40 days. Array CGH analysis showed a pathogenic abnormal result with a 331 Kb heterozygote intragenic deletion into IMMP2L gene. IMMP2L intragenic deletions were detected in 7 cases among 188 unrelated patients with GTS as reported in a previous Danish cohort and one of them present a similar 331 Kb deletion indicating a possible recurrent genomic instability in this region. However, partial deletions of this gene have been observed in 0.9% of the control population in the Danish cohort and in 2 other cases among 2150 CGH array analyses performed in our laboratory and presenting an unrelated phenotype. These data suggest that IMMP2L gene deletions are recurrent finding with variable pathogenic effect and high recurrence of GTS condition.

PrgmNr 2253 - Lipid transporter TMEM24/C2CD2L plays critical roles in ensuring the survival of retinal rod cells

View session detail

Author Block: **Y. Yang**[1,2], X. Zhu[2]; [1]Sch. of Med., Univ. of Electronic Sci. and Technology of China, Chengdu, China, [2]Sichuan Provincial Key Lab. for Human Disease Gene Study, Sichuan Provincial Peopleâ□□s Hosp., Chengdu, China
Disclosure Block: **Y. Yang:** None.

C2CD2L/TMEM24, an endoplasmic reticulum (ER)-localized phospholipid transporter, maintains lipid transport between ER and plasma membrane (PM). The lipid transport function of TMEM24 is essential for sustaining the intracellular Ca^{2+} oscillations, supporting a key role of TMEM24 in the coordination of Ca^{2+} and phosphoinositide signaling. C2CD2L is enriched in neuroendocrine cells and previously suggested to be required for a normal secretory response. However, little is known about the roles of this protein in retina. To this end, we investigated the effect of genetic deletion of *C2cd2l* in retinal rods on the visual function and on pathological changes in mouse. Rho-Cre mice was crossed with *C2cd2l* transgenic mice to yield *C2cd2l* Rod-KO (RKO) progenies. ERG recordings manifested a reduction in the amplitudes of a- and b-waves in RKO mice. Both H&E staining and immunohistochemistry of retinas sections from 4-month-old mice revealed the loss of rod cells and shortened photoreceptor outer segments in RKO retina. Besides, activated astrocytes detected in RKO retinas by immunostaining with GFAP suggested severe glial reaction. Mechanistically, RNA-seq results indicated that the expression levels of several key genes were reduced in RKO retinas, which were involved in multiple biological process, such as voltage-gated calcium channel activity (*Cacna1f, Cacng7, Cacna2d2*), postsynaptic membrane regulation (*Dlg4, Ptprs, Adnp, Cux2, Dnm1*) and visual perception (*Reep6, Rho, Aipl1, Rpgrip1, Fam161a*). These results were further verified by qPCR. Thus, in retinal rods, C2CD2L should participate in these processes to maintain normal visual function. Our data demonstrated novel essential roles of C2CD2L in the retina.

PrgmNr 2254 - Novel variants in *UBE3B* lead marfonoid body habitus of blepharophimosis ptosis intellectual disability syndrome

View session detail

Author Block: **A. Albakheet**, R. Al-Mass, H. AlQudairy, A. Qari, R. Huma1, H. Ghebeh, A-O. Mohammad, D. Colak, N. Kaya, M. Al-Sayed; King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia
Disclosure Block: **A. Albakheet:** None.
UBE3B, an E3 class ligase, is involved in ubiquitination of short-lived proteins and targets them for protein degradation pathway. Recently, defects in this gene are reported to cause Kaufman oculocerebrofacial syndrome. We employed comprehensive genetic analyses including homozygosity mapping, candidate gene sequencing, exome sequencing, and confirmatory Sanger sequencing on a cohort of 139 neurodevelopmental cases and screened the cohort for putative mutations. Here, we report three consanguineous families with 8 patients harboring three novel variants, a missense substitution on HECT-domain in family 1 and a three basepair deletion within exon 14 (leading to removal of a serine) in family 2 and family 3 has a splice site mutation in *UBE3B.* The variants were fully segregated with the phenotype in the tested families members. Iterative filtering of exome sequencing did not reveal any other candidate genes. Further screening on the 500 ethnically matching controls as well as a comprehensive search on local (inclusive of more than 2000 samples), international (public) and commercially available databases confirmed the novelty of three variants. Unlike several other features, blepherophimosis, telecanthus, ptosis and low serum lipid profile were similar to those of previously reported cases having pathogenic *UBE3B* variants. Longitudinal follow up of the selected patients in both families revealed rather tall marfanoid body habitus of our patients. Our study expands phenotypic spectrum of the disease.

PrgmNr 2255 - Phenotypic variability of MEGF10 variants in a consanguineous population

View session detail

Author Block: **H. AlQudairy**[1], M. AlMuhaizea[1], O. Dabbagh[1], W. Alotaibi[1], A. Alhargan[1], M. Mahmoud Ali[1], H. AlHindi[1], D. Colak[1], N. Kaya[2]; [1]KFSHRC, Riyadh, Saudi Arabia, [2]KFSH & RC, Riyadh, Saudi Arabia
Disclosure Block: **H. AlQudairy:** None.
Congenital myopathies are group of inherited neuromuscular disorders with neonatal or infancy onset. Their most common symptoms are hypotonia, muscle weakness and dysmorphic features. One of the rare autosomal recessive congenital myopathy is early onset myopathy, areflexia, respiratory distress, and dysphagia (EMARDD, OMIM: 614399, MIM: 612453). It is caused by biallelic mutations (homozygous or compound heterozygous) in *MEGF10* gene (multiple epidermal growth factor-like domains protein family). There are only five cases reported in the literature to date. Saudi Arabia is a highly consanguineous population and harbor significant number of autosomal recessive rare disorders. Interestingly, no *MEGF10* associated EMARDD has been reported up to date. In this study, we took the initiative and searched our neurology clinics and local databases to identify individuals with this disease. We determined two unrelated consanguineous families with patients having novel *MEGF10* mutations. Our study provides insight into clinical details of the patients and expands phenotypic spectrum of the disease

PrgmNr 2256 - Second report of *SHMT2* related neurodevelopmental disorder with cardiomyopathy, spasticity, and brain abnormalities

View session detail

Author Block: **V. Bhat**[1], P. Majethia[1], B. L. Yatheesha[2], A. Shukla[1]; [1]Kasturba Med. Coll., Manipal, India, [2]Dheemahi Child Neurology and Dev. center, Shimoga, India
Disclosure Block: **V. Bhat:** None.
Mitochondrial one carbon metabolism enzyme serine hydroxymethyltransferase, encoded by *SHMT2* (MIM 138450), plays a crucial role in amino acid and folate metabolism. Biallelic variants in *SHMT2* has been recently described to cause a novel neurodevelopmental disorder with cardiomyopathy, spasticity, and brain abnormalities (NEDCASB; MIM# 619121) in five subjects from four unrelated families. A six-month-old male, first born to third-degree consanguineous parents, presented with developmental delay and occasional jerky movements. On examination at six months of age, his weight was 6kg (-2.5 SD), length was 58cm (-4.9 SD) and head circumference was 37cm (-6.5 SD). He had shallow orbits, anteverted nares, thin vermillion borders, spasticity and exaggerated deep tendon reflexes. Biochemical investigations showed mildly elevated lactate (21mg/dL; ref: 4.5 - 19.8mg/dL). Electroencephalogram showed multifocal epileptiform abnormalities. A magnetic resonance imaging (MRI) of the brain showed thinning of the corpus callosum, particularly at the splenium and hypomyelination. Exome sequencing identified a likely pathogenic novel missense variant c.1133A>G, p.(Asp378Gly) in exon 10 of *SHMT2* (NM_005412.6) in a homozygous state. On segregation, the variant was seen in heterozygous state in the parents. The clinical findings in the reported individuals are subtle dysmorphism, cardiac abnormalities, intellectual disability, peripheral neuropathy, and motor dysfunction in the form of spasticity and ataxia. Cardiac anomalies included hypertrophic cardiomyopathy in four and atrial septal defect in one individual. Clinically, cardiovascular examination of our subject was normal, however complete cardiac evaluation by echocardiography is planned on follow-up visit. Hypertrophic cardiomyopathy has been seen as early as 3 years of age earlier. MRI findings in all previously reported individuals have thin corpus callosum and perisylvian polymicrogyria. The present proband showed thin corpus callosum, particularly at the splenium and delayed myelination. No cortical abnormalities were seen. The electroencephalogram was abnormal in the present subject similar to one of the reported individuals, in absence of clinical seizures. Peripheral neuropathy has not been evaluated in the present subject. Mild lactic acidosis was seen in present proband, however, previously reported individuals have not been evaluated for the same. Till now, a total of seven variants including six missense and one deletion-insertion have been reported in *SHMT2*

PrgmNr 2257 - Simultaneous screening for SMN copy numbers and 2+0 silent carrier genotypes for Spinal Muscular Atrophy

View session detail

Author Block: **P-S. Lai**, G. Y. Chia, C. Liu, P. S. Low, S. K. Tay; Natl. Univ. of Singapore, Singapore, Singapore, Singapore
Disclosure Block: **P. Lai:** None.
Spinal muscular atrophy (SMA) is an autosomal recessive neuromuscular disorder that affects the lower motor neuron cells. It is the most common genetic cause of infant mortality and is caused by homozygous deletions of the *SMN1* gene in ~94% of patients, with remaining cases carrying a heterozygous deletion and a pathogenic variant on each allele. The reported carrier frequency varies between 1 in 35 to 1 in 60 individuals among different populations. Determination of *SMN1* copy number is important for molecular diagnosis and carrier identification. Due to the high incidence and severity of the disease, population-wide SMA screening can identify couples-at-risk and facilitate appropriate genetic counselling. Since carriers are asymptomatic, they can only be detected through *SMN1* copy number screening. However, carrier screening using dosage analysis can present false negative risks for silent carriers with the 2+0 genotype who are not picked up due to having one chromosome carrying two copies of *SMN1* and the other chromosome having 0 copy of *SMN1*. In this study, we analyzed 100 individuals from SMA affected families using the CarrierMaxâ□¢ *SMN1/SMN2* (Thermo Fisher Scientific Inc.) assays to investigate both SMN1/2 copy numbers and 2 SNPs (g.27134T>G and g.27706-27707delAT) associated with the 2+0 genotype. Multiplex qPCR was performed and the amplicons were separated via capillary electrophoresis on the SeqStudioâ□¢ Genetic Analyzer (Applied Biosystems) and analysed on GeneMapper 6.0 and CarrierMax software. A total 12 patients with 0 copy of *SMN1* was observed while family members carrying 1 to 2 or more copy numbers of *SMN1* and *SMN2* were detected. There were 55 carriers with 1 copy of *SMN1*. Validation of the copy numbers was performed by other gene dosage comparisons (MLPA and qPCR). In addition, putative silent carriers were detected (3 with g.27706-27707delAT and 1 with g.27134T>G). No gene conversion event was detected among this cohort. We found this screening approach to be convenient and rapid as diagnosis of copy number and detection of carriers, including silent carriers could be done simultaneously.

PrgmNr 2258 - The suppressive role for the splicing regulator, PTBP1, in the production of neuron-specific Agrin isoform

View session detail

Author Block: **S. Bushra**, Y. Lin, A. Masuda, A. Joudaki, K. Ohno; Nagoya Univ., NAGOYA, Japan
Disclosure Block: **S. Bushra:** None.
Agrin is a ubiquitously expressed proteoglycan, which has diverse functions by tissue-specific isoform expression. The neuron-specific isoforms (neural agrin), but not other isoforms, specifically induce acetylcholine receptor clustering, which is indispensable for the formation of neuromuscular junctions. Neural agrin is generated by the inclusion of three alternative exons called â□□exon Yâ□□, â□□exon Z8â□□ and â□□exon Z11â□□, although the underlying mechanism is not well understood. Here, we investigated splicing *cis*-elements and trans-factors involved in the splicing of exons Y/Zs, and explored their roles in the production of neural agrin. Publicly available CLIP-seq data analysis revealed several trans-factors, such as MATR3, RBFOX, and Polypyrimidine Tract Binding Protein (PTBP) 1, which bind extensively around exons Y/Zs of AGRN pre-mRNA. The knockdown experiments revealed that PTBP1 but not others suppressed the inclusion of exons Y/Zs in the SH-SY5Y neuroblastoma cell line. Consistently, RT-PCR analysis showed that the differentiation of SH-SY5Y cells downregulated PTBP1 expression and increased the inclusion of exons Y/Zs. The bindings of PTBP1 were prominent in the flaking introns of exon Y, which contains 5 polypyrimidine-rich sites. The analysis of AGRN minigenes with mutations of these sites identified two crucial sites, designated PM2 and PM1, responsible for splicing exon Y and exon Zs, respectively. We confirmed that the tethering of PTBP1 to PM2 specifically promoted the inclusion of exon Y. We propose that the neuron-specific splicing of exon Y and Zs in the AGRN gene is facilitated by the suppression of PTBP1 expression.

PrgmNr 2259 - Transcriptomic profiling of chromatin-related neurodevelopmental disorders in human induced pluripotent stem cell -based models

View session detail

Author Block: **R. Woldegebriel**[1], P. Puigdevall[2], Z. Shadman[3,4], O. Arshad[4], E. Metzakopian[3,5], S. Gerety[3,4], M. Hurles[3,4], H. Kilpinen[1,2,3,4]; [1]Helsinki Inst. of Life Sci., Univ. of Helsinki, Helsinki, Finland, [2]UCL Great Ormond Street Inst. of Child Hlth., London, United Kingdom, [3]Open Targets, Wellcome Genome Campus, Cambridge, United Kingdom, [4]Wellcome Sanger Inst., Wellcome Genome Campus, Cambridge, United Kingdom, [5]UK Dementia Res. Inst., Wellcome Genome Campus, Cambridge, United Kingdom

Disclosure Block: **R. Woldegebriel:** None.

Neurodevelopmental disorders (NDDs) are characterized by abnormalities of brain development, most commonly manifesting as intellectual disability or autism spectrum disorders. The genetic architechture of NDDs comprises of highly penetrant de novo mutations in protein-coding genes, which account for up to half of genetically determined NDDs, and frequently affect genes that encode for chromatin-modifying and remodelling proteins. Given their central role in gene regulation, loss-of-function of chromatin-related genes are likely to have widespread effects on the transcriptome, but it is currently not known which cell types during neurodevelopment are most affected by such mutations and when, and whether such effects could be therapeutically reversed. Here, we have applied single-cell gene expression profiling (scRNA-seq) of cortical neuron differentiation in induced pluripotent stem cell (iPSC) models of NDD patients and controls. To model the NDD mutations, we have used both patient-derived iPSCs as well as CRISPR-Cas9-engineered knock-outs (KO) of the corresponding disease genes, and carried out pooled differentiations in order to achieve better comparability of samples through reduced heterogeneity. To identify and characterize the cell types, developmental lineages, and cellular processes most affected in NDDs, the cells were profiled at two timepoints during differentiation: cortical neuron precursors (NPC) and mature cortical neurons. The findings from the single-cell experiments will be compared to bulk transcriptomic data from NPCs derived from a larger number of isogenic iPSC lines with engineered KOs of ~30 known NDD genes, as well as further perturbation experiments with existing chromatin-targeting drugs and a KO screen of 500 chromatin-associated genes, also in NPCs. The aim of the study is to identify novel therapeutic targets for NDDs by comparing transcriptomic and other cellular responses to different types of perturbations across known NDD genes.

PrgmNr 2260 - Ultra rare truncating mutations of GRIK family genes associated with schizophrenia disrupt the interaction with PSD95 protein

View session detail

Author Block: **M-C. Cheng**, T-M. Hu, Y-Y. Wang, C-L. Wu, S-H. Hsu, H-Y. Tsai; Yuli Branch, Taipei Veterans Gen. Hosp., Hualien county, Taiwan
Disclosure Block: **M. Cheng:** None.
Schizophrenia is a complex mental disorder with a high genetic component in its etiology. Identification of the genetic underpinnings of schizophrenia is the key to decipher the pathogenesis of schizophrenia. The ionotropic glutamate receptor kainate type plays a critical role in regulating the synaptic transmission and the functions of various synaptic receptors. GRIK gene family encoding the ionotropic glutamate receptors of kainate subtypes can be considered the candidate genes of schizophrenia. We screened for the rare and pathogenic mutations at the protein-coding sequences of 5 genes (*GRIK1*, *GRIK2*, *GRIK3*, *GRIK4*, and *GRIK5*) in 516 unrelated patients with schizophrenia using the ion semiconductor sequencing method. We conducted the functional assay using immunoblot, immunocytochemistry, and NanoBRETâ□¢ bioluminescent resonance energy transfer (BRET) assay. We identified 44 protein-altering variants, including 12 in *GRIK1*, 5 in *GRIK2*, 7 in *GRIK3*, 13 in *GRIK4*, and 7 in *GRIK5* in patients with schizophrenia and *in silico* analysis showed that some of these mutations were damaging or pathological to the protein function. Notably, we identified two frameshift deletion mutations (GRIK1^{Phe24fs} and GRIK1^{Thr882fs}) in two unrelated patients with schizophrenia and two nonsense mutations (GRIK2^{Arg300Ter} and GRIK4^{Gln342Ter}) in two unrelated patients with schizophrenia. These four truncating mutations have minor allele frequency (MAF) of less than 0.01% from the public database or were absent in 1517 healthy controls from Taiwan BioBank. Functional analysis with immunoblot and immunocytochemistry identified these four truncating mutants as the loss-of-function mutants in HEK-293 cells. BRET assay showed that three truncating mutations (GRIK1^{Phe24fs}, GRIK1^{Thr882fs}, and GRIK2^{Arg300Ter}) weaken the interaction with PSD95 protein. The study suggests that the GRIK genes harbor ultra-rare truncating mutations in certain patients with schizophrenia, supporting that rare coding variants contribute to the genetic architecture of schizophrenia. Future studies using transgenic mouse models carrying these truncating mutations are necessary to understand the functional consequence of the GRIK gene family mutants *in vivo* and, most importantly, to elucidate whether and how they contribute to the etiology of schizophrenia.

PrgmNr 2261 - Whole-genome sequencing analysis identifies rare Alzheimerâ□□s disease risk variants among Chinese individuals

View session detail

Author Block: **J. Cao**[1], Y. Li[1], C-Y. Lo[1], Q. Guo[2], Zhangjiang Internaitional Brain Bank, J. Chen[1], X-M. Zhao[1]; [1]Inst. of Sci. and Technology for Brain-inspired Intelligence, Fudan Univ., Shanghai, China, [2]Dept. of Gerontology, Shanghai Jiao Tong Univ. Affiliated Sixth People's Hosp., Shanghai, China
Disclosure Block: **J. Cao:** None.
Genome-wide association studies (GWAS) have identified more than 30 loci implicated for Alzheimerâ□□s disease (AD), but there is still a substantial proportion of missing heritability which may be explained by rare variants. Here, we performed a whole-genome sequencing analysis on 198 elderly Chinese including 28 individuals with AD, 66 cognitive normal controls and other 104 cognitive declined individuals. Through a pipeline integratively considering allele frequency, functional impact, and pathogenicity, we found 33 rare pathogenic variants exclusively existing in 22 AD individuals but not in normal controls, which affected 29 high confidence AD-associated genes involved in endocytosis, immune response, and amyloid precursor protein (APP) metabolism. Among genes proposed by previous GWAS projects, we discovered 4 variants presenting exclusively in 2 AD samples, including a homozygous missense variant (rs756786529) in *CHRNE* gene, emphasizing the applicability of the cholinergic hypothesis in Alzheimerâ□□s disease. Furthermore, we focused on rare pathogenic variants enriched in AD group (OR > 5), and identified 4 variants that were successfully replicated in 10,913 unrelated case/control samples. We particularly noticed a rare missense variant rs139437968 in *SAMD11* gene. *SAMD11* had a significantly lower expression level in hippocampus from AD brains compared with control brains, and showed a strong negative correlation between its expression and AD pathology in AÎ² line AD mouse models, indicating it can be a possible new biomarker for AD. In summary, our analysis screened multiple rare variants and possible new genes related to AD pathogenesis, highlighting the role of rare variants in AD genetic architecture and their possible indications for diagnostic and therapeutic uses.

PrgmNr 2262 - Microtubule-Associated Serine/Threonine Kinases family genes (MAST) involved in the pathogenesis of seborrheic dermatitis may be related to Parkinson disease

View session detail

Author Block: **A. KARRA**, B. Abdelmoula, N. Bouayed Abdelmoula, Genomics of signalopathies at the service of medicine UR17ES36; Med. Univ. of Sfax, Sfax, Tunisia
Disclosure Block: **A. Karra:** None.
There is a growing body of evidence indicating the implication of abnormal protein kinase function in various aspects of Parkinsonâ□□s disease (PD) etiology and the possible association of microtubule defects and PD pathogenesis. In fact, LRRK2 gene, through deregulation of microtubule assembly, has been demonstrated as the most commonly mutated gene in the clinically typical, late-onset PD as well as both familial and sporadic forms. On the other hand, PD is an entity with a highly increased prevalence of seborrheic dermatitis (SD), which can be seen in approximately 60% of PD patients. SD is a common chronic inflammatory skin disorder affecting 1 to 3 percent of the general population, but even more commonly found in PD patients with a prevalence of 52 to 59 percent. The exact mechanism underlying the association of PD and SD is not fully clear. The increased sebum production associated to the increased reproduction of Malassezia yeasts with a nonspecific immune response are the two major factors believed to contribute to the development of SD. A direct involvement of Malassezia as the common denominator in the pathogenesis of both SD and PD is recently suggested. Moreover, the strong association between SD and PD is supposed to be due to genetic pathways link. Here we present a novel hypothesis concerning a potential molecular mechanism underlying the association of PD and SD that can be associated to the immune-mediated inflammatory process and endorsing the known skin-brain axis theory. In fact, a recent pilot genome-wide association study conducted to identify genetic variants associated with SD, showed two significant SNPs mapped to the MAST4 gene at chromosome 5 and to an intergenic region at chromosome 17p12, between the genes PIRT and SHISAQ. Through a literure review and databases analysis we concluded that microtubule-associated serine/threonine kinases family genes (MAST) involved in the pathogenesis of SD may be related to PD. MAST genes encode proteins members of the microtubule-associated serine/threonine protein kinases family. These proteins contain a domain that gives the kinase the ability to determine its own scaffold to control the effects of their kinase activities. Alternative splicing results in multiple transcript variants encoding different isoforms. MAST kinases named MAST1 to MAST4 are characterised by containing a serine/threonine kinase domain and a postsynaptic density protein-95/discs large/zona occludens-1(PDZ) domain. MAST4 gene is associated with spinocerebellar ataxia 27 and epilepsy with generalized tonic-clonic seizures. It has a ubiquitous expression including the brain and the skin.

PrgmNr 2263 - RNAseq characterization of the effect of titin truncating variants

View session detail

Author Block: **M. Savarese**[1], M. Johari[1], A. Vihola[1], H. Luque[1], M. Iacomino[2], P. Hackman[1], B. Udd[1]; [1]FolkhÃ¤lsan Res. Ctr., Univ. of Helsinki, Helsinki, Finland, [2]IRCCS Istituto Giannina Gaslini, Genoa, Italy

Disclosure Block: **M. Savarese:** None.

Titin related diseases include dominant cardiomyopathies and, mainly recessive, skeletal muscle diseases. Due the sheer size of the *TTN* transcripts, the interpretation of titin variants is challenging. Titin truncating variants (TTNtv) have been associated with a dominant cardiomyopathy with a reduced, age dependent, penetrance. On the other hand, TTNtv carriers do not show any skeletal muscle disease. Congenital skeletal muscle titinopathies are mainly due to bi-allelic TTNtv, with a large proportion of patients having variants affecting the splicing. The ACMG/AMP guidelines for variant interpretation underline that splicing defects are an important mechanism of pathogenicity. To estimate the clinical impact of variants potentially impacting the splicing, we re-evaluated by SpliceAI all the possible single nucleotide changes in the titin coding sequence, identifying 9,732 variants predicted to alter the splicing (threshold 0.5). Of these, 2,078 are synonymous or missense variants and 317 are nonsense variants. Few previously reported causative â□□missenseâ□□ variants are predicted to impact the splicing. In order to characterize the effect of titin truncating variants on skeletal muscle transcripts, we analysed by RNA sequencing 44 skeletal muscle samples from patients with a confirmed or a suspected titinopathy. Our study suggests that variants causing a premature stop codon, located in exons out of the M-band, result in a nonsense mediated decay of titin transcripts. Vice versa, the in-depth characterization of 15 variants affecting canonical splice sites demonstrates that most of them cause in-frame losses or gains, still resulting in a near-full length protein. A direct analysis of RNA, cDNA and protein is crucial to characterize the effect of splice variants. However, further studies are still needed to support the variant classification of DNA changes causing in-frame losses or gains.

PrgmNr 2264 - Automated prediction of the clinical impact of copy number variants: The power of combining expert and machine learning approach

View session detail

Author Block: **J. BudiÅi**[1,2,3], T. SlÃideÄek[1], M. KucharÃk[1,2], M. GaÅ¾iovÃi[1,4], Z. PÃ¶s[1,5,6], O. PÃ¶s[1,5], M. HlavaÄka[1], M. Å tevko[1], W. Krampl[1,2,5], R. Hekel[1,3,5], J. RadvanszkÃ½[1,2,6], T. Szemes[1,2,5]; [1]Geneton Ltd, Bratislava 4, Slovakia, [2]Comenius Univ. Sci. Park, Bratislava, Slovakia, [3]Slovak Ctr. of Scientific and Technical Information, Bratislava, Slovakia, [4]Dept. of Computer Sci., Faculty of Mathematics, Physics and Informatics, Comenius Univ., Bratislava, Slovakia, [5]Dept. of Molecular Biology, Faculty of Natural Sci., Comenius Univ., Bratislava, Slovakia, [6]Inst. of Clinical and Translational Res., BioMed. Res. Ctr., Slovak Academy of Sci., Bratislava, Slovakia

Disclosure Block: **J. BudiÅi:** Salary/Employment; Geneton Ltd.

Introduction: Copy number variants (CNVs) play an important role in many biological processes, including the development of genetic diseases, making them attractive targets for genetic analyses. The interpretation of the effect of structural variants is a challenging problem due to highly variable numbers of gene, regulatory or other genomic elements affected by the CNV. The state-of-the-art scoring scheme proposed by the American Academy of Medical Genetics (ACMG) is a well-respected guideline for the interpretation. The proper evaluation of the scheme is however challenging even for experts skilled in clinical genetics.

Materials and Methods: We automated several steps of the ACMG scoring scheme, proposing clinicians the recommended choices along with enclosed explanatory genomic annotations. In addition, we implemented a novel method based on machine learning that uses its own modeling beyond the ACMG scheme to predict the clinical impact of CNVs.

Results: We demonstrate the high accuracy of the automated scoring of the ACMG scheme and compare it with the accuracy of the machine learning approach. We show that the combination of these two complementary methods accurately predicts the impact of the majority of CNVs extracted from the ClinVar database.

Conclusions: Prediction of the clinical impact of CNVs can be automated, relieving highly valued clinical professionals of tedious annotation and interpretation processes.

PrgmNr 2265 - Coding DNA numbering errors occurring in multiple pipelines

View session detail

Author Block: **S. Seo**, S. Hwang, K. Park; Seoul Natl. Univ. Bundang Hosp., Seongnam, Korea, Republic of
Disclosure Block: **S. Seo:** None.
Introduction: Accurate variant calling in NGS data is critical in clinical sequencing. Herein we report a nucleotide numbering issue that arises from the errors in coding positions of the exons in a specific gene. Methods: A patient with progressive hearing loss underwent a targeted sequencing by using a multigene panel consisted of the genes related to hearing loss. The library was prepared and sequenced on Ion Torrent S5 XL (Thermo Fischer Scientific, USA). Reads were aligned to hg19 and variant calls were made by three different pipelines. Results: All three pipelines detected two variants in *PTPRQ*, chr12:81004405G>A and chr12:81043449G>A, which seemed to be the genetic cause for the patientâ□□s condition. However, all three pipelines showed different cDNA numbering in these two variants (c.4402+1G>A, c.4462+1G>A and c.4464+1G>A for chr12:81004405G>A; c.5508+1G>A, c.5568+1G>A and c.5570+1G>A for chr12:81043449G>A). Accurate nomenclature for these two splice site variants was confirmed as c.4918+1G>A and c.6024+1G>A. The difference were found in coding positions of *PTPRQ* exons defined in each pipeline, which seemed to result in this cDNA numbering error. Conclusions: The coding DNA reference sequences may be updated or replaced by another record regularly, which might result in numbering errors in particular genes. Since accurate nomenclature of a pathogenic variant is critical in a clinical laboratory setting, additional use of a validating tool for variant nomenclature can be of help for the clinical sequencing report.

PrgmNr 2266 - Copy number variation analysis with targeted next generation sequencing in patients with inherited metabolic disorders

View session detail

Author Block: **M. Reboun**, R. Svacinova, V. Todorovova, M. Novakova, D. Musalkova, H. Vlaskova, G. Storkanova, K. Peskova, T. Honzik, L. Dvorakova; Dept. of Paediatrics and Inherited Metabolic Disorders, First Faculty of Med., Charles Univ. and Gen. Univ. Hosp., Prague, Czech Republic
Disclosure Block: **M. Reboun:** None.
Introduction: We implemented a targeted NGS for the diagnosis of patients with a suspicion for various groups of inherited metabolic disorders including glycogen storage diseases, disorders of amino acid/organic acid metabolism and fatty-acid oxidation disorders. **Methods**: A total of 237 genes were analyzed using a custom-designed oligo capture probe set (Roche NimbleGen) and Illumina MiSeq system. Inhouse bioinformatic pipeline was used for of SNPs, small deletions/duplications and copy number variants analysis. Using a CNVkit, copy number variation (CNV) analysis was performed on 24 samples (one sequencing run) with variable sizes of resolution of the deleted/multiplicated regions - â□□binsâ□□: 50bp, 100bp and 150bp. **Results**: In a cohort of 99 patients, in whom the diagnosis was genetically confirmed, ten large deletions were identified. Extensive duplication has not been found yet. Deletions detected in four patients involved only one exon (i.e. deletions of exon 17 in *PYGM*, exon 16 in *HADHB*, exon 1 in *DBT* were present in heterozygous state while exon 33 in *PHKA2* was identified in hemizygous state). Deletions detected in five patients contained two to ten exons of one gene, none of them encompassed the neighboring genes. Thus, deletions including exons 2-3 in *SLC3A1*, 1-10 in *HADHA*, 2-13 in *PHKB*, 12-17 in *CBS*, 10 -12 in *IVD* were detected. The presence of identified CNVs was confirmed by qPCR and/or by Sanger sequencing of PCR products overlapping the deletion boundaries. In the last presented female patient CNV analysis suggested monosomy of X chromosome. This result was confirmed by cytogenetic methods. **Conclusions**: Our results show, that CNV analysis for small panels (i.e. hundreds of genes) and small numbers of samples (i.e. 24) is possible. Analysis of variable bin size showed comparable results, however, 50bp resolution showed a lot of poorly predicted CNVs at the ends of the investigated regions. Resolution of 100-150bp analysis exhibited more precise results and revealed variants at the range of the only one exon to the entire chromosome. In our cohort CNV was detected in 10 (10%) out of 99 diagnosed patients. The presented data indicate, that CNV detection is an important part of mutation analysis. Support: MZ_CR - RVO_VFN64165, SVV260367

PrgmNr 2267 - Evaluating RNA-seq gene expression profiles for disease prediction using machine learning

View session detail

Author Block: **V. Lanke**, R. Srinivasan, S. Datta; Tata Consultancy Services, Hyderabad, India
Disclosure Block: **V. Lanke:** None.
The use of individual genotypes in screening and diagnosis of inherited Mendelian rare disorders is becoming prevalent. While the use of polygenic risk scores to evaluate the likelihood of multigenic disorders is gaining currency, it is understood that augmenting genotypes with intermediary phenotypes such as gene expression and methylation status can lead to improved risk assessment models. Unlike with genotypes, where the presence or absence of a variant required to evaluate risk can be readily reckoned from a reference genome, the lack of an accepted standard reference for gene expression leads to challenges in interpreting individual gene expression profiles. Since there are public repositories of gene expression (SRA, GTEx, TCGA for example) that include a number of healthy controls we hypothesized that we may be able to use these data to create disease specific models by integrating data from these resources with gene expression data from samples with the disease. In this study, we collected data from recount2 - a repository of uniformly processed RNA-Seq samples from multiple resources. Using feature importance metric for iterative feature elimination, we built random forest based disease specific models to distinguish disease and healthy state given an expression profile. Using a test set of four diseases - Chronic Obstructive Pulmonary Disease(COPD), Dilated Cardiomyopathy(DCM), Systemic Lupus Erythematosus(SLE) and Breast Cancer(BC), we demonstrate that we can achieve reliably high classification performance using disease specific models. Although the approach holds promise, the performance of a disease model in the presence of other diseases from the same tissue remains a challenge to be explored.

PrgmNr 2268 - Human Genome Topology for Selected Trios from the 1000 Genomes Project

View session detail

Author Block: **D. M. Plewczynski**[1,2], M. Lazniewski[1,3], M. Chilinski[1,2], S. Gadakh[4], Y. Ruan[5]; [1]Ctr. of New Technologies, Univ. of Warsaw, Warsaw, mazovia, Poland, [2]Faculty of Mathematics and Information Sci., Warsaw Univ. of Technology, Warsaw, Poland, [3]CeZaMaT, Warsaw Univ. of Technology, Warsaw, Poland, [4]Ctr. of New Technology, Warsaw, Poland, [5]The Jackson Lab. for Genomic Med., Farmington, CT
Disclosure Block: **D.M. Plewczynski:** None.
In this work, we present ***a comprehensive analysis of human genome topological changes due to the genomic variability at the population and single family scale*** [Sadowski et al, Genome Biology 2019]. In recent years the number of personal genomes has increased substantially allowing the identification of single nucleotide polymorphisms and structural variants common for the human population. A major challenge is to understand which of these changes have structural or functional implications [Chaisson et al, Nature Comm 2019]. The goal of this work is to understand the link between changes in the DNA sequence, the 3D structure of the genome, and gene expression. In detail the lymphoblastoid cell lines from 9 individuals have been chosen for this study. The samples represent 3 families originating from three geographical locations - China, West Africa, and Puerto Rico.

First, we have used 1000 Genomes public data from Oxford Nanopore sequencing technology along with short-read Illumina, focusing on the detection of the structural variants (SVs) improvement. We improved the quality of the Structural Variants identification from the whole genome sequencing (WGS) experiments by using the consensus approach. Fifteen gold-standard silico callers were used for obtaining the polished list of Structural Variants for each family. The results of the SV callings were merged using our novel ConsensuSV algorithm, which integrates the SV sets using machine learning by combining decision trees and neural networks trained and benchmarked on the high-quality SVs from the Human Genome Structural Variation Consortium. We provide the validated sets of high-confidence Structural Variants identified for each of the analyzed daughters from Trios.

Having identified the SVs dataset from both short-read and long-read techniques, we studied the influence of SNPs and SVs on the genome 3D structure and explored the heritability of genome organization. We performed the differential peak analysis for the Yoruban trio using CTCF ChIA-PET data and identified more than 18 000 unique loci with CTCF binding peak. For 10 336 loci, CTCF peaks were present in all the samples. More than 90% of all loci were also involved in the formation of DNA loops. Next, we incorporated the RNA-seq data to select only regions characterized by an altered gene expression between family members. We presented the mechanism that by changing the specific 3D interactions can increase gene expression, likely by joining promoter and enhancer. We also compare the experimental CTCF ChIA-PET results obtained for the selected trio with our in silico prediction algorithm [Wlasnowolski et. al., NAR 2020].

PrgmNr 2269 - Mining risk regulatory variants of Tetralogy of Fallot using deep learning models

View session detail

Author Block: **Q. Lin**[1], H. Hong[2], P. J. Gruber[3], C. S. Tang[1,4], K. Chien[5], P. K. H. Tam[1,4]; [1]Dept. of Surgery, Li Ka Shing Faculty of Med., The Univ. of Hong Kong, Hong Kong, China, [2]Dept. of Cardiothoracic Surgery, Shanghai Children's Med. Ctr., Shanghai Jiao Tong Univ. Sch. of Med., Shanghai, China, [3]Dept. of Surgery, Yale Univ. Sch. of Med., New Haven, CT, [4]Dr Li Dak-Sum Res. Ctr., The Univ. of Hong Kong-Karolinska Inst.d Collaboration in Regenerative Med., Hong Kong, China, [5]Dept. of Cell and Molecular Biology, Karolinska Inst., Stockholm, Sweden
Disclosure Block: **Q. Lin:** None.
Tetralogy of Fallot (TOF) is the most common form of cyanotic congenital heart defects with a disease prevalence of ~1 in 3,000 newborns. Over the past decade, studies on the human genome and transcriptome have revealed that non-coding variations contribute significantly to human disorders through dysregulating gene expression. However, most of the rare disease-causing non-coding variants were explored using transcriptomic data due to the complexities of gene regulation on the genome level. Recent development of deep learning models has largely improved the performance of alternative splicing and regulatory effect prediction based on sequence context and epigenomic data. Here, we aim to explore if rare noncoding variants contribute to the genetic etiology of TOF through applying deep learning models on whole genome sequencing (WGS) data of 148 TOF trios (patient and unaffected parents) of Chinese ancestry. From the WGS data, we first detected non-coding *de novo* variants (DNVs) and applied SpliceAI and MMSplice to identify those DNVs that are likely to result in alternative splicing and hence loss of function of the genes. Secondly, we applied HeartENN, a neural network-based model trained with 184 human and mouse heart epigenomic features, to predict the regulatory potential of the non-coding DNVs in TOF patients. We observed an increased burden of likely damaging non-coding DNVs (HeartENN scores â¥ 0.1) in TOF patients compared to controls (*P*

PrgmNr 2270 - Reanalysis of population cohort WGS revealed considerable minor allele frequency differences

View session detail

Author Block: **J. Hsu**[1], P-L. Chen[1], C-Y. Chen[1], D-C. Wu[1], S-H. Shih[1], Y-C. Tsai[1], W-A. Chen[2]; [1]Natl. Taiwan Univ., Taipei City, Taiwan, [2]Natl. Taiwan Univ. Hosp., Taipei City, Taiwan
Disclosure Block: **J. Hsu:** None.
Population cohort whole-genome sequence represents the overall genetic variation profile as a strong foundation of implementing precision medicine in the national healthcare system. A comprehensive variants discovery framework is crucial to ensure the accuracy and reproducibility in detail. For a large-scale sequencing project, the systematic noise from the analytic pipeline can introduce a large number of false-positive calls, which will heavily impact the cohort minor allele frequencies, leading the type 1 errors for concluding false associations between genotype and phenotypic trait. Therefore, it is highly recommended to adopt a compatible analytic pipeline for case and control studies, especially compare the cohort allele frequency with the Genome Aggregation Database (gnomAD). We applied and reanalyzed 1,496 Taiwan Biobank (TWB) whole-genome sequence data to represent the comprehensive genetic variation profile of the Taiwanese population. Combining a standard material (NA12878) from Genome In A Bottle Consortium (GIAB) with the corresponding validated results, we benchmarked the variant discovery pipeline for single nucleotide variation (SNV) and small insertion and deletion (INDEL). Variant quality score recalibration (VQSR) suggested by GATK best practice has been used to stratify all variants based on each variant-level quality score. We found more than 4.7% of nonsynonymous variants in the TWB cohort have substantial minor allele frequency (MAF) differences (> 0.05) compared to the East Asian (EAS) population data in gnomAD v2.1. MAF differences may influence a significant proportion of the corresponding phenotypic traits to be considered differently in healthcare settings. Furthermore, there were 0.58% of population-specific variants between TWB and EAS, suggesting the importance of Taiwanese population cohort data for clinical disease diagnosis in Taiwan. As our analysis pipeline has been benchmarked by comparing to GIAB data, the MAF inconsistency should be due to population stratification or small sample size. More population WGS data and a reproducible analysis framework to reanalyze the biobank cohort periodically are necessary to achieve compatible MAF information with other populations.

PrgmNr 2271 - VulExMap: Detection of exons which are vulnerable to exonic splicing mutations

View session detail

Author Block: **L. L. Holm**[1], K. K. Flugt[1], T. K. Doktor[2], M. B. Hansen[1], B. S. Andresen[2]; [1]Univ. of Southern Denmark, Odense M, Denmark, [2]Univ Southern Denmark, Odense M, Denmark
Disclosure Block: **L.L. Holm:** None.
It is now well established that exonic mutations may cause aberrant splicing by altering splicing regulatory elements (SREs) located outside the splice sites. Consequently, software for predicting mutation effects on individual SREs is now widely used in precision medicine. Unfortunately, it is typically presumed that all exons are equally dependent on SREs and that predicted effects of mutations on SREs have the same effect in all exons. However, exonic splicing mutations (ESMs) tend to cluster in particular exons of a gene and similar mutations in SREs located in different exons do not always have identical effects. Furthermore, the existing tools for predicting splicing mutations are predominantly focused on detecting mutations in the splice sites, where a mutation typically may cause aberrant splicing simply by directly affecting splice site strength. We have defined a new subgroup of constitutive exons that are more likely to be affected by ESMs because they have an inherent suboptimal splicing efficiency and therefore are more dependent on the balance between positive and negative SREs. We coin these exons as vulnerable and have created a tool, VulExMap, which uses large RNA-seq datasets to assign exon vulnerability, based on small but significant levels of constitutive exon skipping. We show that approx. 25 % of all constitutive exons are vulnerable, and that vulnerable exons as a group displays characteristics of a suboptimal splicing context, such as weaker splice site strengths, lower density of positive SREs and higher density of negative SREs. VulExMap currently uses 5 different datasets to categorize exon vulnerability and allows the user to upload their own mutations to analyze if they are located in a vulnerable exon. Furthermore, we have generated a database of 202 validated ESMs, and we show by analysis with VulExMap that ESMs are several fold more frequent in vulnerable exons. Moreover, we demonstrate that ESMs that reside in vulnerable exons generally produce smaller changes to the splicing code, whereas ESMs in more resilient exons result in a more severe change to the splicing code. Our goal is that VulExMap will be used as a tool to determine exon vulnerability as a pre-requisite before assessing the potential effects of exonic mutations on splicing. VulExMap is available at https://vulexmap.compbio.sdu.dk

PrgmNr 2272 - A metagenome-wide association study revealed disease-specific landscape of the gut microbiome of systemic lupus erythematosus in Japanese

View session detail

Author Block: **Y. Tomofuji**[1], Y. Maeda[1], E. Oguro-Igashira[1], T. Kishikawa[1], K. Yamamoto[1], K. Sonehara[1], D. Motooka[2], Y. Matsumoto[3], H. Matsuoka[4], M. Yagita[4], N. Takuro[1], S. Oshima[4], S. Nakamura[3], H. Inohara[1], K. Takeda[1], A. Kumanogoh[1], Y. Okada[5]; [1]Osaka Univ. Graduate Sch. of Med., Suita, Japan, [2]Res. Inst. for Microbial Diseases, Osaka Univ., Suita, Japan, [3]Res. Inst. for Microbial Diseases, Suita, Japan, [4]NHO Osaka Minami Med. Ctr., Kawachinagano, Japan, [5]Osaka Univ. Graduate Sch. of Med., Suita, Osaka, Japan
Disclosure Block: **Y. Tomofuji:** None.
Alteration of the gut microbiome has been linked to the pathogenesis of systemic lupus erythematosus (SLE). However, a comprehensive view of the gut microbiome in SLE and its interaction with the host remains to be revealed. Whole metagenome shotgun sequencing technology is useful for evaluating the human gut microbiome and has many advantages over conventional 16S ribosomal RNA (rRNA) sequencing such as higher taxonomic resolution and applicability for functional analysis. However, evaluation of the microbiome-disease association based on shotgun sequencing is still incomplete for SLE, hindering us from understanding the link between the gut microbiome and the etiology of SLE. Furthermore, the insufficient number of shotgun sequencing studies in the non-European population is problematic given the significant impact of ethnicity and lifestyle on the microbial landscape. Here we performed a metagenome-wide association study (MWAS) of SLE based on shotgun sequencing of the gut microbial DNA from 250 Japanese individuals, composed of 47 SLE patients and 203 healthy controls. Our MWAS consisted of three major bioinformatic analytic pipelines (phylogenetic analysis, functional gene analysis, and pathway analysis). Additionally, we integrated the result of the MWAS with the genome-wide association study (GWAS) data and plasma metabolite data for revealing the interaction between the gut microbiome and the host. Via species level phylogenetic analysis, we identified and validated increases of *Streptococcus intermedius* and *Streptococcus anginosus* in the SLE patients. Microbial gene analysis revealed increases of *Streptococcus* derived genes including one involved in the redox reaction. Additionally, microbial pathways related to sulfur metabolism and flagella assembly were altered in the SLE patients. We identified a pathway level SLE-specific link between the metagenome and the germline genome by comparing the result of the current MWAS and GWAS of SLE (i.e., MWAS-GWAS interaction). Î±- and Î²-diversity analyses provided evidence of dysbiosis in the metagenome of the SLE patients. Microbiome-metabolome association analysis identified a positive dosage correlation of acylcarnitine with *Streptococcus intermedius*, an SLE-associated taxon. Thus, our MWAS followed by integrative analysis with the GWAS and the metabolite data revealed SLE-associated changes in the gut microbiome and its interaction with the host. Our analysis provided novel insights into the pathogenesis of SLE and could be a key step toward the implementation of more appropriate therapeutic frameworks.

PrgmNr 2273 - A multi-disciplinary approach to solving undiagnosed patients - Unsolved Cases Unit Groningen

View session detail

Author Block: **D. Bos**[1], C. M. A. Van Ravenswaaij-Arts[1], T. Dijkhuizen[1], A. H. van der Hout[1], K. Kok[1], K. Van Dijk-Bos[1], L. Zijlstra[1], D. S. Verbeek[1], F. De Andrade[1], R. Pfundt[2], N. Corsten-Janssen[1], B. Sikkema-Raddatz[1], M. Van Gijn[1], K. M. Abbott[1]; [1]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [2]Radboud Univ. Nijmegen Med. Ctr., Nijmegen, Gelderland, Netherlands
Disclosure Block: **D. Bos:** None.
Introduction: After solving 30-40% of all patients submitted for clinical genetic testing, the possible avenues for follow-up testing are diagnostically limited. Experimental diagnostics can help to bridge the gap between the clinic and research, providing the next step after diagnostic genetics has been exhausted. At the UMCG, we have implemented a multi-disciplinary team focused on solving well-phenotyped unsolved cases using an experimental diagnostic approach.
Methods: A team of clinicians, researchers, laboratory geneticists and a patient representative was established. A secure portal was created for registering unsolved cases. A trial-run of the unit was initiated to assess its inclusion criteria, accessibility, feasibility, as well as its added value. Subsequent tests included mRNA and WGS, novel bioinformatics tools for prioritizing variants and functional testing. The progress was tightly monitored; biweekly meetings were organized to monitor progress and regular updates were shared with clinicians to avoid endless experimental diagnostic tracks. Over a six month period, 10 out of 17 patients registered were chosen for inclusion. The phenotypic spectrum included developmental delay, MCA, skin ailments and suspected hereditary cancer. Five patients required further functional testing to enable re-classification for VUS. The remaining five were chosen for whole genome sequencing.
Results: To date, a homozygous missense variant in *PLAA* was found to initiate alternative splicing using mRNA techniques, resulting in truncation of the gene. With this result the phenotype of the patient, who has brain anomalies, hypotonia, edema of the hands, episodes with bradycardia and saturation loss, epilepsy and feeding difficulties, is explained. A VUS in *TBCK* (compound heterozygous with pathogenic variant) in a patient with muscular hypotonia, seizures, motor delay, severe intellectual disability was not found to have an effect on splicing (using a mini-gene assay technique). Further mRNA testing is being performed in connection with another study. A VUS missense variant in *EDA* in family members with ectodermal dysplasia is currently being tested. Using RT-PCR and Western blot techniques we will compare *EDA* function in one healthy/VUS-neg family member with two affected/VUS-pos family members. Testing for the remaining patients is ongoing.
Conclusions: With the invaluable support of the department, the assistance of a patient representative and an enthusiastic team, the Unsolved Cases Unit - Groningen was piloted. Preliminary results indicate the added value of such a unit in genome diagnostics to assist in solving the unsolved cases.

PrgmNr 2274 - An atlas of genetic scores to predict multi-omic biomolecular traits in blood

View session detail

Author Block: **Y. Xu**[1], S. C. Ritchie[1], M. Pietzner[1], S. Lambert[1], S. May-Wilson[2], A. Nath[3], P. Surendran[1], Ã. Johansson[4], E. Persyn[1], L. Lannelongue[1], B. Prins[1], N. Pirastu[5], D. Paul[1], C. Yau[6], J. F. Wilson[7], C. Langenberg[1], A. MÃ¤larstig[8], J. Danesh[1], A. S. Butterworth[1], M. Inouye[1]; [1]Univ. of Cambridge, Cambridge, United Kingdom, [2]Univ. of Edinburgh, Edinburgh, United Kingdom, [3]Baker Heart and Diabetes Inst., Melbourne, Australia, [4]Uppsala Univ., Uppsala, Sweden, [5]Univ. of Edinburgh, Edinburgh, Midlothian, United Kingdom, [6]Univ. of Manchester, Manchester, United Kingdom, [7]Univ Edinburgh, Edinburgh, United Kingdom, [8]Karolinska Inst.t, Stockholm, Sweden
Disclosure Block: **Y. Xu:** None.
Introduction: Genetically predicted levels of biomolecular traits have been shown to be a useful tool for investigating the molecular underpinnings of complex phenotypes. However, existing studies mainly focus on molecular traits of a single or a small number of biological domains (e.g. gene and protein expression) sourced from a variety of different studies. There is currently a lack of resources that leverage the large-scale multi-modal data of single cohorts, thus facilitating well-powered and cross-modality analyses in cohorts where multi-omic data are unavailable but genotype data are.
Objective: Develop an open resource of genetic scores for transcriptomic, proteomic and metabolomic traits, to enable any genotyped cohort to generate accurate and comparable predictions of multi-modal omics data.
Method: We utilized the INTERVAL study, a 50,000-participant cohort of blood donors with extensive multi-omic profiling, including genome-wide genotypes and biomolecular traits for plasma proteomics (SomaScan, and Olink), serum metabolomics (Nightingale), plasma metabolomics (Metabolon), and gene expression (whole-blood RNA sequencing). We trained genetic score models for these >24,000 biomolecular traits on the INTERVAL cohort utilizing a machine learning method and evaluated the performances of the genetic scores within the cohort and externally in four other cohorts. Furthermore, we applied the genetic scores to impute biomolecular traits into UK Biobank and subsequently performed a phenome-wide association scan (PheWAS) for predicted biomolecular trait associations.
Results: First, we illustrate the consistent performance of the constructed genetic score models between the internal and external validation, and that many traits have highly predictive scores (e.g. 98 SomaScan proteins, 21 Olink proteins and 6 Metabolon metabolites had an $r^2 > 0.3$ in both validations). We also illustrate the utility of our scores in a PheWAS, where we replicate well-known blood proteins/metabolites to disease associations, e.g. blood C-reactive protein and total cholesterol with coronary artery disease (CAD), and discover new associations where no previous molecular data was available.
Conclusions: We present a resource for genetic imputation of biomolecular traits that can boost cross-cohort, cross-domain analyses, and enable large-scale data integration for greater powered studies to identify biomarkers, therapeutic targets and pathways.

PrgmNr 2275 - Enzymatic DNA synthesis (EDS) enables rapid and broad-based access to synthetic oligos needed for the genetic analysis and functional characterization of the SARS-CoV-2 virus

View session detail

Author Block: **B. Derrien**, S. Romero; DNA Script, SAS, Le Kremlin-BicÃªtre,, France
Disclosure Block: **B. Derrien:** None.
The COVID-19 pandemic sparked a global scientific effort to study the epidemiology, genetics, biochemistry and evolution of the SARS-CoV-2 virus. One unforeseen effect of this work was a worldwide bottleneck in the supply of synthetic DNA (primers, probes, assay controls, and gene fragments), which currently relies on highly centralized phosphoramidite-based production and third-party logistics.
We have developed a novel enzymatic DNA synthesis (EDS) technology that enables decentralized, same-day, on-demand oligo production in a standard molecular biology lab. Here we report the use of EDS primers in (i) amplicon sequencing and phylogenetic analysis of clinical SARS-CoV-2 samples and (ii) site-directed mutagenesis of the SARS-CoV-2 spike (S) gene.
For SARS-CoV-2 amplicon sequencing, the ARTIC v3 panel (218 primers) was produced by EDS. Libraries generated from two synthetic RNA controls and five Mexican isolates yielded coverage uniformity and variant calling data similar to that obtained with commercial primers. Phylogenetic analysis of clinical samples showed no correlation with isolates from countries where patients had travelled prior to testing positive.
In another study, the unique ability of EDS to extend existing ssDNA was used to produce a custom version of the ARTIC v3 panel, with 5â□□-overhangs to add Illumina adaptors in a second PCR. By eliminating ligation-based library prep, the sample-to-sequencing turnaround time was reduced to a single day. This simplified protocol enabled high-throughput monitoring of SARS-CoV-2 variants in Finland.
EDS primers were also used to produce an expression plasmid encoding the spike SARS-CoV-2 Alpha/B.1.1.7 variant for functional studies. Oligos were designed to introduce ten mutations into a cloned copy of the 3.8-kb S gene using a multi-site approach. Sanger sequencing primers produced by EDS were used to confirm individual mutations in picked clones and tiled sequencing of the entire S gene. EDS primers showed high success rates in a multi-site mutagenesis strategy and yielded high-quality Sanger sequencing data. In-house oligo synthesis enabled a fast, iterative process that needs only a few days per mutagenesis/sequencing cycle.
Data generated in collaboration with The Jackson Laboratory for Genomic Medicine, USA; the University of Zacatecas Molecular Medicine Laboratory and Zacatecas Public Health Laboratory, Mexico; the Institute for Molecular Medicine, Finland; and the Pasteur Institute, France.

PrgmNr 2276 - Impaired type I interferon activity of subjective cognitive decline in preclinical Alzheimerâ□□s disease

View session detail

Author Block: **L. Song**[1], J. Chen[1], C-Y. Lo[1], Q. Guo[2], Zhangjiang Internaitional Brain Bank, X-M. Zhao[1]; [1]Fudan Univ., Shanghai, China, [2]Shanghai Jiao Tong Univ. Affiliated Sixth People's Hosp., Shanghai, China

Disclosure Block: **L. Song:** None.

Subjective cognitive decline (SCD) is considered to be the preclinical syndrome of Alzheimerâ□□s disease (AD), which is a potentially crucial window for preventing or delaying the progression of this disease. No previous study has evaluated the transcriptomic profilings on SCD in preclinical AD. To explore the etiology of AD and identify candidate biomarkers, we comprehensively assessed the peripheral blood transcriptomic disruption for SCD, including lncRNA, mRNA, and miRNA. Dysregulated protein-coding mRNA at gene and isoform-level and integrative coexpression network consistently implicated an impaired type I interferon signaling pathway in SCD. The pathway was regulated by differential splicing of interferon genes, *NRIR*, and *has-miR-146a-5*, hub genes, and transcription factors. Normal cognition individuals with lower expression of hub genes *STAT1* and *TRIM22* exhibited a higher MCI and AD conversion rate in the ADNI cohort, indicating that *STAT1* and *TRIM22* might serve as candidate biomarkers for disease progression. Our findings demonstrate a down-regulation of type I interferon activity in preclinical AD, which would increase the risk of disease progression.

PrgmNr 2277 - Molecular Cartography: a multiplexed high resolution transcriptomics approach to spatially analyze rare events demonstrated in an infection model of SARS-CoV-2

View session detail

Author Block: **S. Groiss**[1], D. Pabst[1], C. Faber[2], A. Meier[2], A. Bogdoll[2], C. Unger[2], B. Nilges[2], E. FÃ¶derl-HÃ¶benreich[1], M. Hardt[1], A. Geipel[2], F. Reinecke[2], C. Korfhage[2], K. Zatloukal[1]; [1]Med. Univ. of Graz, Graz, Austria, [2]Resolve BioSci.s GmbH, Monheim am Rhein, Germany
Disclosure Block: **S. Groiss:** None.
COVID-19, the clinical manifestation of the SARS-CoV-2 infection, exhibits severe heterogeneity in infection rates and cellular responses in different cell types. Such heterogeneity necessitates the molecular analysis of single cell gene expression in a highly multiplexed manner. Many regulatory genes responsible for these processes are expressed in low copy numbers and the cellular and subcellular location impacts their function. Single-cell spatial transcriptomics technologies allow investigations of intricate processes in cells such as host-to-pathogen responses, antiviral defense pathways and mapping the response dynamics in different infection sites. The methods and technologies available thus far lack a direct comparison of multiple samples and are limited in detecting rare events due to reduced sensitivity, specificity, or resolution. To gain a thorough comparative understanding of the infection rate of various cell lines by SARS-CoV-2 and their molecular response, we used a multiplexed single-cell mRNA in-situ hybridization technique developed by Resolve Biosciences called Molecular Cartographyâ¢ (MC) that allows for the detection and quantification of rare transcripts at a subcellular level. In this study, we demonstrate that MC detects single mRNA species at a specificity of > 99 % and a sensitivity comparable to single-molecule FISH that is considered the gold standard of spatial in-situ methods. We illustrate the utility of MC in a SARS-CoV-2 infection model using cell lines of pulmonary, colorectal, and hepatocellular origin. Using MC, we localized transcripts of the SARS-CoV-2 nucleocapsid protein majorly vicinal to the cell nuclei, highlighted cell-to-cell variances in expression levels of the key viral entry factors *ACE2*, *TMPRSS2* and *FURIN*, and revealed virus-induced dysregulation of major primary and secondary antiviral genes. By spatially tracking antiviral responses emanating from the primary infected cell, we elucidate differences in individual, temporal divergent infection sites and trace individual regulatory signatures in pathways hijacked by SARS-CoV-2. Overall, MC is demonstrated to be a novel, multiplexed single-molecule spatial technology to explore complex cellular responses at unprecedented specificity, sensitivity, and resolution.

PrgmNr 2278 - Translational diagnostics program -TDP-, an innovative intramural approach for hospital-based functional genomics of undiagnosed and rare diseases

View session detail

Author Block: **F. Palau**[1,2,3,4], J. Pijuan[2], M. Rodriguez-Sanz[2], A. Altisent-Huguet[2], M. Roldan[1], J. Armstrong[1], D. Yubero[1], J. Maynou[1], L. Cantarero[3], G. Fernandez[1], J. Hoenicka[2]; [1]Hosp. Sant Joan de DÃ©u, Barcelona, Spain, [2]Inst. de Recerca Sant Joan de DÃ©u, Barcelona, Spain, [3]CIBER de Enfermedades Raras (CIBERER), ISCIII, Barcelona, Spain, [4]Hosp. ClÃnic and Univ. of Barcelona, Barcelona, Spain

Disclosure Block: **F. Palau:** None.

INTRODUCTION: For the best management and treatment of patients with undiagnosed and rare diseases (URDs), it is necessary to innovate in the diagnostic process. In a significant number of patients, genome sequencing has increased the rate of diagnosis of URDs, but it also detects genetic variants of unknown clinical significance (VUS) or inconsistent with the phenotype. To assist physicians in the variant classification concerning patient's phenotype we have developed the in-house Translational Diagnostics Program (TDP)[1]. OBJECTIVES: To validate for diagnosis variants that are classified as a VUS or novel phenotype-genotype correlation using the holistic approach â□□precision phenotyping - clinical genomics - functional genomics - team decision-makingâ□□. RESULTS: Between 2017 and 2020, we conducted 4,149 clinical exome sequencing studies at the Department of Genetic Medicine at SJD Children's Hospital. The genetic diagnosis was achieved in 1,360 patients (33%) and in 717 (17%) we detected genetic variants classified as VUS following the ACMG/AMP standards and guidelines. To address functional studies of a VUS or a phenotype-genotype incongruity we applied the TDP pipeline that includes: (1) a comprehensive evaluation of the phenotype, including HPOs; (2) *in silico* analysis of the pathogenicity of the candidate genetic variant/s; (3) functional validation of the variant by examine the encoded protein using molecular and cellular assays, and comparative computational analysis of confocal microscopy images; (4) diagnostic decision-making with referring physicians. Currently, 51 patients have been included in the TDP for biological validation of the candidate genetic variant/s. Most patients are affected by neurological diseases such as neuromuscular, neurodevelopmental (mainly syndromic), or epileptic disorders. We confirmed in 27 out of 51 patients a deleterious impact of the VUS upon the location/function of the encoded protein. Of the remaining patients, in 5 cases the variant had no functional impact on the coded product, 13 are under study, and 5 are still under discussion. CONCLUSIONS: The genetic diagnosis of URD patients requires the promotion and implementation of intramural functional genomics/diagnostics programs to support and resolve medical problems when the patient's variant is uncertain or inconsistent with the phenotype. The TDP is a strong in-house tool to fill the gap between clinical phenotype and genotype. [1]Pijuan et al. J Mol Diagn 2021, doi:10.1016/j.jmoldx.2020.10.006. Support by: ISCIII; Generalitat de Catalunya and FEDER; FundaciÃ³n Isabel Gemio; TorrÃ³ Solidari RAC1 i Torrons Vicens; and IRSJD - Carmen de Torres.

PrgmNr 2279 - An automated workflow for high-throughput qPCR

View session detail

Author Block: **S. Tomei**, H. Shobha Manjunath, M. Kumar Reddy Kalikiri; Sidra Med. and Res. Ctr., Doha, Qatar
Disclosure Block: **S. Tomei:** None.
Automation solutions can significantly improve sample processing efficiency and can be of particular help in core facility settings. Centralized core facilities are being established world-wide with the aim at strengthening institutionsâ clinical and research enterprises and at addressing the need to process large volumes of samples on expensive cutting-edge technologies in a limited time. High-throughput qPCR profiling is a service offered by most genomics facilities. Several platforms have been developed to process large numbers of samples in a short time, including the Fluidigm Biomark HD. Several automation systems are currently available to miniaturise volumes and improve bioanalytical workflows, including the SPT Labtech Mosquito HV system. Here we applied the Mosquito HV platform for the automation of the sample preparation of the Fluidigm gene expression workflow. We successfully automated the pre-amplification and exonuclease cleanup steps with the aim at reducing manual error and sample processing time. We show consistency in the expression of house-keeping genes when assessing pooled RNA control samples for the manual and automated workflow of Fluidigm gene expression profiling.

PrgmNr 2280 - An integrated pipeline for genome analysis and phenotype-based gene prioritization with GenDiseak

View session detail

Author Block: **J-H. Huang**[1], T-F. Chen[1], Y-H. Lin[1], J-Y. Yu[1], P-H. Li[1], Y-C. Tu[1], H-K. Tsai[1,2], H-F. Juan[1,3,4], C-Y. Chen[1,5]; [1]Taiwan AI Labs, Taipei, Taiwan, [2]Inst. of Information Sci., Academia Sinica, Taipei, Taiwan, [3]Dept. of Life Sci., Natl. Taiwan Univ., Taipei, Taiwan, [4]Graduate Inst. of BioMed. Electronics and Bioinformatics, Natl. Taiwan Univ., Taipei, Taiwan, [5]Dept. of Biomechatronics Engineering, Natl. Taiwan Univ., Taipei, Taiwan

Disclosure Block: **J. Huang:** None.

Genetic examination through next-generation sequencing has been integrated into routine clinical practice in the modern era. However, to interpret the most relevant genetic variations through whole genome or exome sequencing remains a significant challenge for most clinical utilities in terms of bioinformatics analysis capacity. To accelerate the process of molecular diagnostics and reduce the turnaround time, we developed GenDiseak, a cloud-based genetic disease diagnosis platform. GenDiseak provides an end-to-end (from fastq to causative variants) solution for the application of clinical genetics on a single sample and family-based samples. GenDiseak has built multiple variant analysis and interpretation strategies for diseases with a comprehensive phenotype-gene association from different resources. First, we curated the widely-used databases, including ClinVar and PanelAPP, which help rapidly identify causal genes associated with particular phenotypes. Second, we employed state-of-the-art deep learning techniques in natural language processing to annotate the associations between phenotypes and genes on the same sentences from the full text of medical literature using a Transformer model automatically. We further used dNorm to normalize phenotype entities into the Medical Subject Headings (MeSH) ID in order to match the gene-phenotype pairs. Therefore, users could also input free text of phenotypic descriptions for patients without diagnosed diseases to prioritize a gene list based on their genome sequencing data. In summary, GenDiseak is an innovative tool for the automatic genome analysis and prioritization of phenotype-based gene panels. It thus facilitates and accelerates the variant interpretation in clinical genetics practices.

PrgmNr 2281 - Implementation of heart disease-associated cellular heterogeneity in vivo with Gene-Disease Integrative Systems Transgenesis (GENISYST) and Molecular Cartography

View session detail

Author Block: **N. Kashikar**[1], B. Nilges[1], J. Maxeiner[2], D. Stefanoska[2], J. Kaur[3], A. Bogdoll[1], J. Krishnan[3,2]; [1]Resolve BioSci. GmbH, Monheim am Rhein, Germany, [2]Genome Biologics, Kronberg im Taunus, Germany, [3]Inst. for Cardiovascular Regeneration, Ctr. for Molecular Med., Goethe Univ., Frankfurt am Main, Germany
Disclosure Block: **N. Kashikar:** Salary/Employment; Resolve BioSciences GmbH.
Spatial tissue population heterogeneity in disease arises due to cellular evolution culminating in the establishment of cellular subpopulations with molecular and phenotypic variations - leading to differential drug and therapeutic responses. Although best established in the cancer paradigm, the impact of disease heterogeneity in non-neoplastic disease has similarly been observed in cardiovascular, metabolic and neurodegenerative indications. However, current transgenic methodologies are not able to facilitate the establishment of complex heterogeneity patterns in vivo. This inability to mimic human disease heterogeneity in animal models is a stumbling block for precision pre-clinical disease modeling in target and therapeutics validation. To overcome these limitations, we have developed a novel methodology enabling the implementation of complex disease heterogeneity patterns through combinatorial transduction of modular genetic units to drive gene-specific loss-of-function (LOF). In analyzing heterogeneity patterns in human diseased myocardium, we identified LOF of Hif1a, Atp5a1, Syt17, Smg1, Mthfd1 and Mthfd1l in distinct subpopulations of cardiomyocytes. We utilized GENISYST to implement these heterogeneity patterns in mouse myocardium and validated them with Molecular Cartography (MC) technology. MC is a novel imaging-based combinatorial single-molecule fluorescent in situ hybridization technology. We detected and quantified the expression of 100 mRNA species at subcellular resolution at high sensitivity and specificity in different layers of the mouse heart with MC. Analysis of hundreds of single mRNA molecules per cell yielded detailed information on the spatial distribution of cardiomyocyte subtypes and their transcriptomic states in the context of their location in the tissue and relative to other cell types. The high sensitivity allowed for the detection of underrepresented transcripts and their cell specificity; potentially serving them as useful drug targets for cell-specific therapeutics. In summary, MC accurately defined the spatial heterogeneity patterning imposed on cardiac subpopulations by GENISYST in native myocardium for the first time.

PrgmNr 2282 - scDetect: A rank-based ensemble learning algorithm for cell type identification of single-cell RNA sequencing in cancer

View session detail

Author Block: **Y. Shen**; Zhejiang Univ., Hangzhou, China
Disclosure Block: **Y. Shen:** None.
Motivation Single-cell RNA sequencing (scRNA-seq) has enabled the characterization of different cell types in many tissues and tumor samples. Cell type identification is essential for single-cell RNA profiling, currently transforming the life sciences. Often, this is achieved by searching for combinations of genes that have previously been implicated as being cell-type specific, an approach that is not quantitative and does not explicitly take advantage of other scRNA-seq studies. Batch effects and different data platforms greatly decrease the predictive performance in inter-laboratory and different data type validation. **Results** Here, we present a new ensemble learning method named as â□□scDetectâ□□ that combines gene expression rank-based analysis and a majority vote ensemble machine-learning probability-based prediction method capable of highly accurate classification of cells based on scRNA-seq data by different sequencing platforms. Because of tumor heterogeneity, in order to accurately predict tumor cells in the single cell RNA-seq data, we have also incorporated cell copy number variation consensus clustering and epithelial score in the classification. We applied scDetect to scRNA-seq data from pancreatic tissue, mononuclear cells, and tumor biopsies cells and show that scDetect classified individual cells with high accuracy and better than other publicly available tools. **Availability** scDetect is an open source software. Source code and test data is freely available from Github (https://github.com/IVDgenomicslab/scDetect/) and Zenodo (https://zenodo.org/record/4764132#.YKCOlrH5AYN). The examples and tutorial page is at https://ivdgenomicslab.github.io/scDetect-Introduction/.

PrgmNr 2283 - Allelic and genotypic frequencies of *CYP2C19*2, CYP2C19*3,* and *CYP2C19*17* alleles in the healthy South Indian population and their phenotypic prediction

View session detail

Author Block: **A. Prashant**, A. Yadav, A. Kodakkat, A. Helna, P. Vishwanath, S. Nataraj; Ctr. of Excellence in Molecular Biology & Regenerative Med., Dept. of Biochemistry, JSS Med. Coll., JSS Academy of Higher Ed. and Res., Mysuru, India
Disclosure Block: **A. Prashant:** None.
Introduction: Pharmacogenomics and pharmacogenetics of *CYP2C19* help to select appropriate medication to avoid toxicity and to determine proper drug dosage regimen based on an individual's genotype. The allelic study of *CYP2C19* plays a vital role in drug therapy and research and it is important to investigate the implications of various alleles on drug metabolism to avoid adverse drug reactions. This study measures the allelic frequencies of *CYP2C19* (*2, *3 and *17) alleles in the general population. Methodology: The study involved 300 healthy subjects between the age group of 15-85 years recruited by simple random sampling from areas in and around Mysuru, South India. Allele-specific touch-down PCR was employed for genotyping. The allelic and genotypic frequencies were calculated and were checked for Hardy-Weinberg equilibrium, the phenotypic prediction of ultra-rapid metabolizer (UM = *17/*17), extensive metabolizer (EM = *1/*17, *1/*1), intermediate metabolizer (IM = *1/*2, *1/*3, *2/*17) and poor metabolizer (PM = *2/*2, *2/*3, *3/*3) was made based on their genotype. Results: The allele frequency of *CYP2C19*2, CYP2C19*3,* and *CYP2C19*17* was 0.365, 0.0033, and 0.18 respectively. The IM phenotype predominated with an overall frequency of 46.67% including 101 subjects with *1/*1, 2 subjects with *1/*3, and 37 subjects with *2/*17 genotype. This was followed by EM phenotype with an overall frequency of 35% including 35 subjects with *1/*17 and 70 subjects with *1/*1 genotype. PM phenotype had an overall frequency of 12.67% including 38 subjects with *2/*2 genotype and UM phenotype had an overall frequency of 5.67% including 17 subjects with *17/*17 genotype. Conclusion: In view of the high frequency of PM in the South Indian population a pre-treatment test to identify the genotype of the individual may be recommended to decide on the dosage and monitor the drug response and avoid adverse drug reactions.

PrgmNr 2284 - Assessment of cytochrome P450 polymorphism for personalized therapy in acute coronary syndrome

View session detail

Author Block: **R. Frikha**, H. Ghozzi; Univ. Hosp., Sfax, Tunisia
Disclosure Block: **R. Frikha:** None.
The cytochrome P450 (CYP) 2C19 isoenzyme plays an important role in clopidogrel metabolization. CYP2C19*17, the c.C806T, is a recently reported variant causing ultra-rapid metabolism of CYP2C19 substrates, which may lead to an enhanced platelet response to clopidogrel treatment with an increased risk of bleeding. Molecular diagnostic testing for this polymorphism is widespread and needs to be standardized. 13 patients with acute coronary syndrome undergoing coronary intervention after pretreatment with clopidogrel were enrolled in this study. Genotyping was performed with a specific PCR-RFLP. 2 patients were heterozygous (CT806) and 1 patient was homozygous (CC806) for variant C806T of CYP2C19*17 with an overall frequency of 23%. A particular history of bleeding risk and high on-treatment reactivity to clopidogrel were recorded in these patients. Genotyping of CYP2C19*17 variant, the c.C806T, prior to clopidogrel treatment is likely to be useful in order to personalize therapy.

PrgmNr 2285 - Association between HLA-DRB1*15:01 and DQA1*01:02 frequencies and Lumiracoxib-induced DILI in healthy Thai Population

View session detail

Author Block: **P. Chokpitakkul**[1], P. Satapornpong[2]; [1]Shrewsbury Intl. Sch. Riverside Bangkok, Bangkok, Thailand, [2]The division of general pharmacy practice, Dept. of pharmaceutical care, Coll. of Pharmacy, Rangsit Univ., Pathum Thani, Thailand
Disclosure Block: **P. Chokpitakkul:** None.
Introduction:Lumiracoxib is used mainly for symptomatic treatment of osteoarthritis and acute pain in patients. However, lumiracoxib is responsible for drug induced liver injury (DILI) and can cause serious life threatening outcomes. We found the genetic markers predictive of lumiracoxib-related hepatotoxicity in Europeans, consisting of HLA-DRB1*15:01 (P=6.8x10-25, OR 7.5 and 95% CI = 5.0-11.3) and HLA-DQA1*01:02 from genome-wide studies. Objectives: The aim of this study was to evaluate the distribution of HLA-DRB1*15:01 and HLA-DQA1*01:02 associated with pharmacogenetics markers of lumiracoxib-induced DILI for screening in Thai population.Materials and Methods:200 participants were recruited from a healthy Thai population. HLA class II alleles were genotyped using sequence-specific oligonucleotides. Results:Our data revealed 41 alleles of the HLA-DRB1*15:01 (10.25%) and 98 alleles of the HLA-DQA1*01:02 (24.50%) in the Thai population. We founded the most allele frequencies are HLA-DRB1*15:02 (15.25%), -12:02 (14.25%), -15:01 (10.25%), -09:01 (8.75%), and -16:02 (7.50%). Furthermore, HLA-DQA1*01:01 (27.75%), -01:02 (24.50%), -03:02 (13.00%), -06:01 (10.25%), and -02:01 (4.50%) in Thais. Conclusion:To conclude, instead of completely withdrawing Lumiracoxib from treatments, it is advisory to screen for the pharmacogenetics markers of HLA-DRB1*15:01 and HLA-DQA1*01:02 alleles in Thai patients for the purpose of avoiding DILI.

PrgmNr 2286 - Association of a single nucleotide polymorphism rs12496846 in the *C3orf20* gene region, identified to be associated with postoperative analgesia after mandibular sagittal split ramus osteotomy, with postoperative analgesia after major abdominal surgery

View session detail

Author Block: **D. Nishizawa**[1], M. Nagashima[2], S. Kasai[1], J. Hasegawa[1], K. Nakayama[1], Y. Ebata[1], M. Hayashida[1,3,4], K. Ikeda[1]; [1]Tokyo Metropolitan Inst. of Med. Sci., Tokyo, Japan, [2]Toho Univ. Sakura Med. Ctr., Sakura, Japan, [3]Saitama Med. Univ. Intl. Med. Ctr., Saitama, Japan, [4]Juntendo Univ. Sch. of Med., Tokyo, Japan
Disclosure Block: **D. Nishizawa:** None.
Purpose: Opioids are commonly used as effective analgesics for the treatment of acute and chronic pain. However, considerable individual differences have been widely observed in sensitivity to opioids. To explore polymorphisms responsible for the inter-subject difference in the opiate sensitivity, we performed an association analysis focusing on rs12496846 single-nucleotide polymorphism (SNP) in the *C3orf20* gene region, which was identified to be associated with postoperative analgesia after mandibular sagittal split ramus osteotomy, to investigate an association of this SNP with postoperative analgesia after major abdominal surgery. **Methods**: The subjects in the association study were a total of 112 patients with written informed consent who underwent major open abdominal surgery in hospitals and were treated with analgesics including opioids after surgery. The study protocol was approved by the Institutional Review Board at each related institute. Total genomic DNA was extracted from peripheral blood or oral mucosa samples by standard procedures and used for genotyping. TaqMan(R) SNP Genotyping Assays was conducted for genotyping the rs12496846 SNP. Further, effect of the SNP on the *C3orf20* gene expression was investigated with RNA and DNA samples offered by Stanley Foundation Brain Bank. **Results**: In an association study in patients undergoing major open abdominal surgery, carrying G alleles in the rs12496846 SNP was found to be significantly associated with increased postoperative 24-h analgesic requirements, as was observed in the previous study in patients undergoing mandibular sagittal split ramus osteotomy (pC3orf20 gene (pConclusions: The results indicate that this SNP could serve as a marker that predict increased analgesic requirements. Our findings will provide valuable information for achieving satisfactory pain control and open new avenues for personalized pain treatment.

PrgmNr 2287 - Comprehensive analysis of the genetics of clozapine-induced adverse effects

View session detail

Author Block: **J. Partanen**[1], M. LÃ¤hteenvuo[2], H. Taipale[2,3], A. Hellsten[4], M. J. Daly[5,1], A. Palotie[5,1], S. Ripatti[1], J. Koskela[1], FinnGen; [1]Inst. for Molecular Med. Finland FIMM, Helsinki, Finland, [2]Dept. of Forensic Psychiatry, Univ. of Eastern Finland, Niuvanniemi Hosp., Kuopio, Finland, [3]Dept. of Clinical NeuroSci., Div. of Insurance Med., Karolinska Inst.t, Stockholm, Sweden, [4]Univ. of Helsinki, Helsinki, Finland, [5]Massachusetts Gen. Hosp., Boston, MA
Disclosure Block: **J. Partanen:** None.
Clozapine is the most effective antipsychotic drug and the only drug available for treatment-resistant schizophrenia (SCZ). Clozapine discontinuation is avoided because of the lack of effective alternatives and potential attenuation of effect at reinitiation. Several adverse effects, some of which are life-threatening, substantially hinder clozapine use, as around 20% of patients discontinue treatment due to them. The genetics of these adverse effects remain largely undiscovered as only agranulocytosis and myocarditis have been studied genome-wide.

We studied the genetics of clozapine-induced adverse effects at scale in the FinnGen biobank study (n = 321,300), including the SUPER cohort of 8,833 patients diagnosed with psychotic illnesses. First, we identified adverse effects by exploring which of 4,131 endpoints were enriched in 2,659 SCZ patients with clozapine purchases compared to 3,628 SCZ patients without, using logistic regression adjusting for age and sex and limiting to incident cases after clozapine initiation or SCZ diagnosis. Second, we studied the genetics of these adverse effects using GWAS comparing incident cases in clozapine users to controls in long-term clozapine users using SAIGE adjusting for age, sex and ten first principal components.

Enriched endpoints (FDR-adjusted p Preliminary GWAS suggested a genome-wide significant association for pneumonia after clozapine initiation at chr13 (intergenic rs1925750(A>G), AF = 66%, p = 4.6E-8, OR = 0.68) in proximity to *PCDH9* previously associated with drug response and a number of psychiatric phenotypes. Further, at chr4 intergenic rs11727255(C>T) (AF = 19%, p = 7.5E-7, OR = 0.66) was associated with adverse drug events and drug allergies in the UKBB (p = 2.6E-4, OR = 0.79).

While GWAS in other endpoints analyzed did not reach genome-wide significance, they included a number of suggestive signals in high LD with Finnish enriched coding variants, in genes involved in drug metabolism, and near previously reported drug-related phenotype associations. All GWAS will be updated to the newest data release.

Our study elucidates trajectories of clozapine treatment and aims to identify genetic determinants of adverse effects related to clozapine, facilitating the progress toward safer use of the most effective antipsychotic in the future.

PrgmNr 2288 - CYP2D6 genetic variants and their metabolic efficacy - insight from Molecular Dynamics Simulations

View session detail

Author Block: **D. Kotzampasi**[1], A. Kanterakis[2], V. Daskalakis[3], G. Potamias[2]; [1]Sch. of Med., Univ. of Crete, Heraklion Crete, Greece, [2]Fndn. for Res. and Technology - Hellas, Heraklion Crete, Greece, [3]Cyprus Univ. of Technology, Limassol, Cyprus
Disclosure Block: **D. Kotzampasi:** None.
Cytochrome P450s is an enzyme superfamily of hemoglobin, responsible for metabolizing more than 90% of clinical drugs. One of the most significant enzymes in this family, Cytochrome P450 2D6 (CYP2D6), metabolizes approximately 25% of the clinically used drugs including crucial and commonly administered drugs such as antidepressants, chemotherapeutics, beta-blockers and opioids. Variations in CYP2D6, a highly polymorphic loci in the genome, could alter its activity influencing the efficacy and toxicity of numerous drugs. More than 100 haplotypes (star alleles) of the drug metabolizing enzyme CYP2D6 have been reported in the Pharmacogene Variation Consortium (PharmVar, www.pharmvar.org), resulting in wide intraindividual variability in drug metabolism activity and changes of the drug plasma concentration. The complete connecting link between the genetic variants and the metabolizer phenotype is still an open and challenging question. Our main objective was to investigate the key factors that determine the metabolizer phenotype by exploiting and appropriately employing molecular dynamics (MD) methods. MD is an elaborate computational method that enables the prediction of the time evolution of atomic positions within interacting systems of molecules. To this end, we have probed the dynamics of numerous CYP2D6 variants, as enzyme models with normal and no function, at an all-atom resolution. With this approach we aimed at filling-in the gaps of missing information and provide a series of analyses, crucial for the prediction of the metabolizer phenotype of CYP2D6. Results are of great importance for areas like Personalized Medicine, Adverse Drug Reaction (ADR) prediction and drug discovery.

PrgmNr 2289 - Distribution of *HLA-B*13:01* allele related with dapsone-induced severe cutaneous adverse reaction in Thai and Asian population

View session detail

Author Block: **M. Leelakajonjit**[1], A. Srisodsai[2], P. Satapornpong[3]; [1]Satit Pattana secondary school, Bangkok, Thailand, [2]MedCoach Inst., Bangkok, Thailand, Bangkok, Thailand, [3]The division of general pharmacy practice, Dept. of pharmaceutical care, Coll. of Pharmacy, Rangsit Univ., Pathum thani, Thailand

Disclosure Block: **M. Leelakajonjit:** None.

Background:Dapsone is antibiotic and anti-inflammatory which is widely used for treatment such as leprosy. However, dapsone cause severe cutaneous adverse reactions (SCARs) include Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN) and drug rash with eosinophilia and systemic symptoms (DRESS), approximately 0.5-3.6% of patients treated with dapsone and 9.9% mortality rate. From the previous studies, only *HLA-B*13:01* allele has a strongly association with dapsone-induced SCARs in Asian population. Moreover, the distribution of *HLA* alleles that play important roles in predicting adverse drug reactions in each population.Objective:Thus, this study was to investigate the distribution of *HLA-B*13:01* allele in Thai and Asian population and importance of this pharmacogenetics marker.Materials and Methods:We recruited 200 unrelated healthy Thai individuals in this study. *HLA-B* were genotyped using sequence-specific oligonucleotides (PCR-SSOs).Results:We found *HLA-B* alleles frequencies in Thai population consist of *HLA-B*46:01* (11.75%), *HLA-B*15:02* (9.25%), *HLA-B*13:01* (6.25%), *HLA-B*4001* (6.25%), and *HLA-B*38:02* (5.50%). For *HLA-B*, *HLA-B*46:01* was the predominant allele commonly found in Thais. This study showed that the frequency of *HLA-B*13:01* allele was similar to the previous study in Thai population. Many publications presented varying distributions of *HLA-B*13:01* in Asians including 9.15% of Han Chinese, 6.67% of Japanese and 5.94% of Vietnamese.Conclusion:Therefore, database of pharmacogenomics containing distribution of *HLA-B*13:01* alleles will support the screening of dapsone-induced SCARs in Thai and Asian population.

PrgmNr 2290 - Distribution of *HLA-B*58:01* allele associated with allopurinol-induced SCARs in healthy Thai population

View session detail

Author Block: **S. Jaroenjiengchai**[1], A. Srisodsai[2], P. Satapornpong[3]; [1]Bodindecha (Sing Singheseni), Bangkok, Thailand, [2]MedCoach Inst., Bangkok, Thailand, [3]The division of general pharmacy practice, Dept. of pharmaceutical care, Coll. of Pharmacy, Rangsit Univ., Pathum Thani, Thailand
Disclosure Block: **S. Jaroenjiengchai:** None.
Background: Allopurinol is the most commonly used for treatment in patients with gouty arthritis, hyperuricemia and cancer patients undergoing chemotherapy. However, allopurinol is a major cause of severe cutaneous adverse reactions (SCARs) in Europeans, Asians, and Thais. We found that the *HLA-B*58:01* is associated with allopurinol-induced SCARs including Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN), and drug reaction with eosinophilia and systemic symptoms (DRESS) in many populations. Furthermore, the frequency of *HLA-B*58:01* allele has interested for the screening among different ethnicities before initiation of treatment.Objective: To study the frequency of *HLA-B*58:01* with Allopurinol-induced SCARs in Thai population.Materials and Methods: 200 general Thai population who had no history of drug-induced SCARs were included in this study. *HLA-B* alleles were genotyped using polymerase chain reaction-sequence-specific oligonucleotides (PCR-SSOs).Results: Among all 200 subjects in Thais, the top five of *HLA-B* alleles consisted of *HLA-B*46:01* (16.00%), *HLA-B*40:01* (8.00%), *HLA-B*15:02* (7.25%), *HLA-B*13:01* (6.50%), and *HLA-B*58:01* (6.25%). Moreover, *HLA-B*58:01* allele was the highest allele in the Asian population approximately 7.38%. On the contrary, we found this allele approximately 1.13% of Caucasians, 1.07% of Hispanics, and 0.80% of North American. Interestingly, *HLA-B*58:01* was the main allele in Thailand and there was found more common in African Americans (6.37%).Conclusion: Our pharmacogenomics database could be used as a screening test for other populations before treatment with allopurinol.

PrgmNr 2291 - Experience and expectations of pharmacogenetic tests in France

View session detail

Author Block: **S. Verdez**[1], Y. Duffourd[1], M. Luu[1], C. Binquet[1], C. Peyron[2], A. Schmitt[2], C. Thauvin-Robinet[3], N. Picard[4], L. Faivre[5]; [1]Dijon Bourgogne Univ. Hosp., Dijon, France, [2]UniversitÃ© de Bourgogne, Dijon, France, [3]FHU-TRANSLAD, Dijon, Dijon, France, [4]CHU Limoges, Limoges, France, [5]Hosp d' Enfants, Dijon, France
Disclosure Block: **S. Verdez:** None.
Although French genomic medicine is reaching a turning point in its history and the implementation of genome sequencing in routine will be possible by the France genomic medicine 2025 plan (PFMG), many questions remain unanswered. The management of secondary pharmacogenetic information is one of them. Therefore, it is necessary to ascertain the opinion of French healthcare stakeholders on the subject.
We created a 29-question questionnaire on the experiences, attitudes, expectations and knowledge of French-speaking physicians and pharmacists towards pharmacogenetics. We compared the responses in different groups and also determined a knowledge score. We created a prediction model for this score and we determined which factors may influence. All differences between these groups were evaluated using Ï□2 and all factors using multivariate logistic regression.
In our study, 90.4% (n=314) of responders thought PGx tests could be a tool to optimize a patient's drug therapy in the future. Therefore, a large majority of responders (90.4%, n=262) were favorable to the incorporation of PGx recommendations into health computerized systems. Moreover, 63.0% (n=335) thought that pharmacogenetic data should be communicated along with the results of the primary analysis on PFMG results when a pharmacogenetic result could influence a therapeutic prescription. Only 9.7% of responders reached the maximum score on the knowlege test. In our regression model, physicians have a lower chance to have a high score (odd ratio: 0.081, pDuring this work we were able to compare the results of a French study with work carried out in other countries. American, Dutch, Canadian and Flemish studies also show similar results. Similarly, Flemish pharmacists achieve higher knowledge scores compared to Flemish physicians.[1] Education is requested by physicians on the topic.
1. Edris A, Vanoverschelde A, Bushaj P, Van Nieuwerburgh F, Lahousse L. Pharmacogenetics in clinical practice: current level of knowledge among Flemish physicians and pharmacists. Pharmacogenomics J. fÃ©vr 2021;21(1):78-84.

PrgmNr 2292 - Integration of genetically regulated gene expression and pharmacological library provides therapeutic drug candidates

View session detail

Author Block: **T. Konuma**[1], K. Ogawa[2], Y. Okada[3]; [1]Central Pharmaceutical Res. Inst., JAPAN TOBACCO INC., Takatsuki, Osaka, Japan, [2]Osaka Univ. Graduate Sch. of Med., Suita, Japan, [3]Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., Suita, Osaka, Japan
Disclosure Block: **T. Konuma:** None.
Since recent studies have shown that drug development projects whose drug targets were supported by human genome information improved these success probability, human genomics-led approaches for drug development have been anticipated. Genomic loci associated with human complex diseases identified by genome-wide association studies (GWASs) are promising genomic information for insights of new therapeutics. Transcriptome-wide association study (TWAS), which predict genetically-regulated gene expression (GREx) by integrating GWAS and expression quantitative trait loci, is one of powerful methods for yielding functional insights into disease susceptibility. We have hypothesized that compounds which have inverse effects in gene expression profiles when compared to GREx from common diseases could be detected as potential drug candidates effective for disease treatment. Here, we constructed the GWAS-TWAS-compound library integration pipeline software (Trans-Phar; integration of Transcriptome-wide association study and pharmacological database) (Konuma T. et al., *Human Molelucar Genetics* 2021). This software conducts *in silico* screening of the compounds from a large-scale pharmacological database, the CMap L1000 library, which have an inverse correlation with GREx estimated from the inputted GWAS summary statistics. Trans-Phar incorporates thirteen tissue or cell-type categories, which we assigned based on the 29 Genotype-Tissue Expression (GTEx) tissues used in TWAS and 77 cell types from the CMap L1000 library database, so that Trans-Phar can detect compounds on a broad scale that affect tissue or cell-type category-specific manners. We applied Trans-Phar to GWAS summary statistics of large-scale European meta-analysis (17 traits; $n_{average}$ = 201,849). As top-associated compounds, anisomycin (FDR-q = 0.056), followed by proscillaridin (Na^+/K^+-ATPase inhibitor) (FDR-q = 0.138) and digoxin (Na^+/K^+-ATPase inhibitor) (FDR-q = 0.138) in central nervous system category in schizophrenia were detected and mechanism of actions of these compounds are likely involved in pathophysiology of schizophrenia. In summary, Trans-Phar successfully identified promising drug target candidates with implications on cell-type-specific pathophysiology of the diseases.

PrgmNr 2293 - Pharmacogenomics Marker of *HLA-A*33:01* Allele Frequency in Healthy Thais

View session detail

Author Block: **N. Chitthiang**[1], A. Srisodsai[2], P. Satapornpong[3]; [1]Singapore Intl. Sch. of Bangkok, Bangkok, Thailand, [2]MedCoach Inst., Bangkok, Thailand, [3]Div. of general pharmacy practice, Dept. of pharmaceutical care, Coll. of Pharmacy, Rangsit Univ., Pathum Thani, Thailand
Disclosure Block: **N. Chitthiang:** None.
IntroductionTerbinafine is an antifungal medication. However, terbinafine has been associated with drug induced liver injury (DILI). Genome-wide association study had shown significant correlations between terbinafine-induced DILI, and HLA-A*33:01 gene (OR=2.6, 95% CI=1.8-3.7, and p-value = 7.0Ã□10â□□8) in Europeans. Interestingly, the distribution of pharmacogenetics markers in each population might differ. ObjectiveThis study aims to investigate the distribution of HLA-A*33:01 gene related to terbinafine-induced DILI in the healthy Thai population. Materials and Methods200 healthy Thais were enrolled in this study who have lived in the area for more than three generations. HLA class I alleles were genotyped by using polymerase chain reaction-sequence specific oligonucleotides (PCR-SSOs). ResultsA total of 33 *HLA-A* alleles were found. Ranked by their frequencies, top 10 were HLA-A*11:01 (27.50%), HLA-A*24:02 (11.50%), HLA-A*02:03 (11.00%), HLA-A*33:03 (10.75%), HLA-A*02:07 (7.50%), HLA-A*02:01 (4.75%), HLA-A*24:07 (4.50%), HLA-A*01:01 and HLA-A*30:01 (2.75%), HLA-A*11:02 and HLA-A*24:10 (2.00%), and HLA-A*02:06 (1.75%). HLA-A*33:01â□□s frequency was 0.25% in the healthy Thai population, while it was 2.95% of Israel, 1.92% of Han Chinese, 1.8% of Brazil, 1.7% of Columbia, 1.1% of Germany, and 0.6% of India. ConclusionThe frequency of HLA-A*33:01 could be used for pharmacogenetics screening before initiation of terbinafine treatment, in order to avoid terbinafine-induced DILI.

PrgmNr 2294 - Practical guidelines of genomics-driven drug discovery from Global Biobank Meta-analysis Initiative

View session detail

Author Block: **S. Namba**[1], T. Konuma[1,2], Y. Okada[1]; [1]Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., Suita, Osaka, Japan, [2]Central Pharmaceutical Res. Inst., JAPAN TOBACCO INC., Takatsuki, Osaka, Japan
Disclosure Block: **S. Namba:** None.
Efficient screening of novel therapeutic targets is an essential process of drug discovery. Disease risk genes by GWAS are promising resources for *in silico* screening. Disease risk genes are known to be enriched in the targets of the drugs approved for the treatment of the diseases themselves, which motivates genomics-driven drug repositioning. However, there does not exist practical guidelines of how to conduct genomics-driven drug discovery, especially for the large-scale GWAS meta-analysis of multiple ancestries. Here, we introduce practical guidelines, as lessons from cross-population GWAS meta-analysis results of a set of common diseases from Global Biobank Meta-analysis Initiative (GBMI). Our framework consists of two parts. (i) Overlap enrichment of disease risk genes with targets of existing drugs. We translate variant-based GWAS p-values into gene-based p-values (e.g., MAGMA). After functional prioritization of biological genes (e.g., DEPICT & Pi), we quantified enrichment of disease risk genes with drug targets, according to disease classification such as WHO ATC and ICD10 codes, using GREP (Sakaue S. *Bioinformatics* 2019). (ii) Integrative approach to screen negative correlations between genetically-determined and compound-regulated case-control gene expression profiles (e.g., tissue-specific TWAS results and the CMAP L1000 database, respectively), using Trans-Phar (Konuma T. *Hum Mol Genet* 2021). We applied our framework to the GBMI GWAS results of 14 phenotypes, which provided a list of candidate drugs for repositioning. In particular, we observed drug target enrichment in asthma, gout, venous thromboembolism (VTE) GWAS. Biological gene prioritization can improve tissue-specificity enrichment and overlap with drug targets, while the choice of prioritization methodologies can induce trends not equally among tissue and drug indication categories. Per-variant effective sample sizes vary strikingly among genome-wide variants in cross-population meta-analysis (only 21.3% of the variants satisfied >50% of effective sample sizes compared to the total ones). We found that this issue can confound biological gene prioritizations, and recommend to restrict the variants to those with more effective sample sizes than the pre-defined threshold. Our work provides the best practices for genomics-driven drug discovery in the era of global biobank collaboration.

PrgmNr 2295 - The frequency of *HLA-B*57:01* allele in healthy Thai population

View session detail

Author Block: **P. Samphatcharoen**[1], A. Srisodsai[2], P. Satapornpong[3]; [1]Chiangmai Univ. Demonstration Sch., Chiangmai, Thailand, [2]MedCoach Inst., Bangkok, Thailand, [3]3.The division of general pharmacy practice, Dept. of pharmaceutical care, Coll. of Pharmacy, Rangsit Univ., Pathum Thani, Thailand

Disclosure Block: **P. Samphatcharoen:** None.

IntroductionsAbacavir hypersensitivity reactions (AHR) are a life-threatening drug reaction and develop in approximately 5-8% of patients receiving abacavir and usually occur between the 6 weeks after initiation treatment. Previous studies showed that the *HLA-B*57:01* allele is strongly associated with abacavir hypersensitivity reactions. Objective This study aimed to investigate the frequency of *HLA-B*57:01* in the healthy Thai population. MethodsWe enrolled 200 healthy Thais and *HLA-B* alleles were genotyped with sequence-specific oligonucleotides (PCR-SSOs).ResultsThis study we found a total of 53 alleles for *HLA-B* genotyping. The most of *HLA-B* alleles in Thais including of HLA-B*46:01 (14.25%), HLA-B*58:01 (6.5%), HLA-B*40:01 (6.25%), *HLA-B*13:01* (5.5%), *HLA-B*44:03* (4.5%), *HLA-B*15:02* (4.5%), *HLA-B*38:02* (4.5%), *HLA-B*44:06* (3.75%), *HLA-B*51:01* (3.75%) and *HLA-B*52:01* (3.25%). Additionally, the allele distribution of *HLA-B*57:01* was 1.75% in Thai population. On the contrary, *HLA-B*57:01* allele was found more commonly in 5.00% of Costa Rica, 4.70% of Chile, 3.15% of East Croatia and 3.30% of China.ConclusionTherefore, the distribution of *HLA-B*57:01* was pharmacogenomics marker, which could be used a screening AHR for all ethnicities.

PrgmNr 2296 - Association between relative telomere length in umbilical cord tissue and PCB congeners in maternal blood and cord blood in a birth cohort in Chiba, Japan

View session detail

Author Block: **T. Takahashi**[1], A. Eguchi[2], M. Watanabe[2], E. Todaka[3], K. Sakurai[4], C. Mori[2,5]; [1]Dept. of Environmental Preventive Med. (Yamada Bee Company, Inc.), Ctr. for Preventive Med. Sci., Chiba Univ., Chiba, Japan, [2]Dept. of Sustainable Hlth.Sci., Ctr. for Preventive Med. Sci., Chiba Univ., Chiba, Japan, [3]Dept. of Global Preventive Med., Ctr. for Preventive Med. Sci., Chiba Univ., Chiba, Japan, [4]Dept. of Nutrition and Metabolic Med., Ctr. for Preventive Med. Sci., Chiba Univ., Chiba, Japan, [5]Dept. of Bioenvironmental Med., Graduate Sch. of Med., Chiba Univ., Chiba, Japan
Disclosure Block: **T. Takahashi:** Grant/Contracted Research Support (External); Yamada Bee Company, Inc..
According to the Developmental Origins of Health and Disease hypothesis, fetal exposure to environmental pollutants during pregnancy increases the risk of non-communicable diseases such as obesity, diabetes, and mental disorder in later life. Environmental exposure to persistent organic pollutants such as polychlorinated biphenyls (PCBs) during pregnancy has been reported to affect the health of the fetuses adversely. Telomere is a highly conserved repetitive DNA sequence. It is currently considered a biomarker for aging. As the telomere has been reported to be affected by exposure to chemical substances, stress, and diseases, it is possible to be used as a biomarker to evaluate health risks by fetal exposure to environmental pollutants. To elucidate the effects of fetal exposure to PCBs, we examined the relationship between maternal and cord blood PCB levels and the umbilical cord telomere length (TL). Our study participants comprised 114 mother-child pairs who had participated in the Chiba Study of Mother and Child Health birth cohort. Maternal blood during pregnancy, cord blood, and umbilical cords at birth were collected at a hospital in Chiba Prefecture, Japan. We extracted genomic DNA from the umbilical cord tissue to measure the TL. Quantitative PCR was used to assess the TL for quantitatively measure the repeat sequence. TL is defined as the ratio of the quantity of the amplicon for the telomere region to the single-copy region of the human Î²-globin (hbg) gene, and is expressed by the formula TL=Telomere / hbg. Serum concentrations of PCB congeners were analyzed using gas chromatography-quadrupole mass spectrometry. We detected 16 and 5 PCB congeners in maternal and cord serum, respectively. We found an association between cord serum PCB levels and TL in the cord tissues of the male fetuses, and the association was significant than between maternal serum PCB levels and TL in the cord tissues. It was also found that among the five congeners detected in cord serum, only PCB74 showed a significant positive association with TL in the cord tissues of the male fetuses. These associations were not found in the female fetuses. As TL could be a biomarker to screen PCB74 exposed male fetuses, further investigation regarding the relationship between PCBs and TL, health effects of PCB74, and the mechanism of the effect difference between the two sexes should be studied.

PrgmNr 2297 - Distinct roles of ADAM10 and ADAM17 for testicular germ cell apoptosis

View session detail

Author Block: **O. Aydos**[1], Y. Yukselten[2], D. Aydos[3], A. Sunguroglu[1], K. Aydos[4]; [1]Ankara Univ. Sch. of Med. Dept. of Med. Biology, Ankara, Turkey, [2]Res. Lab. for Hlth.Sci., Y Gen Biotechnology Company Ltd., Ankara, Turkey, [3]Ankara Univ., Stem Cell Inst., Ankara, Turkey, [4]Ankara Univ. Sch. of Med. Dept. of Urology, Ankara, Turkey
Disclosure Block: **O. Aydos:** None.
Introduction: Germ cells apoptosis during spermatogenesis is a key process to control germ cell maturation and eliminate damaged or unwanted cells. An increase or decrease in the rate of apoptosis can cause defective consequences and lead to azoospermia. The ADAMs, family of transmembrane proteins play an important role in different processes such as cell death, differentiation and migration. ADAM17 and ADAM10 are two members of this family, involved in germ cell apoptosis during spermatogenesis. In this study, our aim was to investigate the role of *ADAM10* and *ADAM17* genes in non-obstructive azoospermia (NOA) etiology. **Materials and methods:** Fold changes in expression levels of *ADAM10* and *ADAM17* were determined by quantitative PCR in NOA (n = 10) patients and obstructive azoospermia (OA) (n = 2) as control cases. Testicular tissue samples were taken during microTESE attempt and testicular biopsy was performed for histopathological evaluation. **Results:** Quantitative PCR results showed a significant decrease and increase in mRNA levels of *ADAM10* and *ADAM17* by 0.23Â± 0.11 (pConclusion: Activation of *ADAM10* could release the extracellular domain of FasL and hinders its binding to Fas in germ cells. Shedding of the extracellular domain of c-kit by *ADAM17* inhibits survival signals of germ cells, leading them to apoptosis. In our study, low expression of *ADAM10* and high expression of *ADAM17* suggest that the germ cells may be directed to apoptosis. This is important for understanding the etiology of NOA. Further studies in larger groups, including protein analysis, are required to clarify the role of *ADAM10* and *ADAM17* in male fertility.

PrgmNr 2298 - Exploring the factors affecting classification and reporting uncertain CMA findings, using a 'virtual fetus' model

View session detail

Author Block: **I. Maya**[1], Y. Yaron[2], A. Reches[2], D. Brabbing-Goldstein[1], R. Michaelson-Cohen[3], H. Yonath[4], M. Regev[4], H. Shani[4], L. Basel-Salmon[1], G. Altarescu[5], R. Segel[5], R. Sukenik-Halevy[6], L. Salzer[1], H. Daum[7], T. Harel[7], V. Meiner[7]; [1]Recanati Genetic Inst., Petah Tikva, Israel, [2]Ichilov Med. Ctr., Tel Aviv, Israel, [3]Shaare Tzedek Med. Ctr., Tel Aviv, Israel, [4]Sheba Genetic Inst., Ramat Gan, Israel, [5]Shaare Tzedek Genetic Inst., Tel Aviv, Israel, [6]Meir Genetic Inst., Netaniya, Israel, [7]Hadassah Genetic Inst., Jerusalem, Israel

Disclosure Block: **I. Maya:** None.

Background: Despite frequent updates in guidelines, classification of copy number variants (CNVs) by different clinical laboratories is inconsistent. Studies exploring the factors affecting clinical decision making are rare, especially in the prenatal setting. The objective of our study was to shed light on these factors, using a 'virtual fetus' model. Methods: Ten prenatally diagnosed CNVs of unknown significance larger than 1 Mb, encompassing OMIM-morbid genes and inherited from a healthy parent (five duplications and five deletions), were independently classified by 15 medical geneticists from five genetic centers. In each referral center, three senior MD specialists were selected, all experienced in CNV classification, representing clinical laboratory, prenatal medicine, and preimplantation genetic testing (PGT) units. Specific points addressed included independent CNV classification, the obligation to report to the parents, recommendation for invasive testing or PGT in future pregnancies, and the relevance of various clinical factors (phenotype and gender of family members carrying CNV, patient request to report any finding, ethnic background, obstetric history, and maternal age) for classification and reporting. Results: The agreement upon classification of various CNVs ranged between 57% to 100%; and was higher for duplications vs. deletions (87% vs. 72%, p=0.013). 84% of the geneticists would report the findings to the parents (92% vs. 76% for deletions and duplications, respectively, p

PrgmNr 2299 - Genome-wide association analysis implicates dysregulation of Wnt signaling pathway in biliary atresia

View session detail

Author Block: **C. S. Tang**[1,2], W. Y. Lam[1,2], Q. Lin[1], M. M. Garcia-BarcelÃ³[1], P. K. H. Tam[1,2]; [1]Dept. of Surgery, Li Ka Shing Faculty of Med., The Univ. of Hong Kong, Hong Kong SAR, China, [2]Dr Li Dak-Sum Res. Ctr., The Univ. of Hong Kong-Karolinska Inst.d Collaboration in Regenerative Med., Hong Kong SAR, China

Disclosure Block: **C.S. Tang:** None.

Biliary Atresia (BA) is a severe, complex hepatobiliary disorder and is a major cause of neonatal cholestasis. It is characterized by progressive fibrosclerosing and inflammatory obliteration of the biliary system in response to unknown insult during the first few weeks of life. Recent genetic researches revealed several BA-susceptibility loci with common risk alleles predisposing to BA; however, the genetic basis of BA and its underlying disease mechanisms remain largely unexplored. With the aim of identifying novel BA-associated variants, we conducted a meta-analysis of genome-wide association studies (GWAS), including 489 non-syndromic BA cases and 1044 controls of East Asian ancestry. We uncovered two BA-associated loci with low frequency variants marginally associated with BA (*USP34*: $P=2.9x10^{-7}$; OR=3.02 [95% confidence interval(CI): 1.98-4.60] and *PTPRZ1*: $P=7.8x10^{-7}$; OR=2.12 [95% CI: 1.57-2.86]). Both *USP34* and *PTPRZ1* play important roles in Wnt/Î²-catenin signalling and were reported to be dysregulated upon liver injury. Specially, BA patients with liver transplant have three times higher risk allele frequency for the top variant at *USP34* (~11.7%) than controls and the frequency is also higher than those BA patients without transplant. Our results showed that Wnt/Î²-catenin signalling may underlie both the aetiology and progression of BA.

PrgmNr 2300 - Genomic sequencing in critically ill newborn infants shows high diagnostic rate in neurometabolic phenotypes and positive impact on clinical management

View session detail

Author Block: **H. Suzuki**[1], M. Nozaki[2], H. Yoshihashi[3], K. Imagawa[4], D. Kajikawa[5], Y. Yamaguchi[6], M. Yamada[1], N. Morisada[7], M. Eguchi[8], S. Ohashi[9], S. Ninomiya[10], T. Seto[11], T. Tokutomi[12], M. Hida[1], K. Toyoshima[13], M. Kondo[14], A. Inui[15], R. Kosaki[16], K. Kurosawa[17], Y. Ito[18], N. Okamoto[19], K. Kosaki[1], T. Takenouchi[1]; [1]Keio Univ., Shinjuku, Tokyo, Japan, [2]Osaka Women's and Children's Hosp., Izumi, Osaka, Japan, [3]Tolyo Metropolitan Children's Hos, Fuchu, Tokyo, Japan, [4]Univ. of Tsukuba, Tsukuba, Ibaraki, Japan, [5]Ibaraki Childrenâ□□s Hosp., Mito, Ibaraki, Japan, [6]Gunma Children's Med. Ctr., Shibukawa, Gunma, Japan, [7]Hyogo Prefectural Children's Hosp., Kobe, Japan, [8]Tokyo Metropolitan Bokutoh Hosp., Sumida, Tokyo, Japan, [9]Tokyo Metropolitan Ohtsuka Hosp., Toshima, Tokyo, Japan, [10]Kurashiki central Hosp., Kurashiki, Japan, [11]Osaka City Univ. Graduate Sch. of Med., Osaka, Osaka, Japan, [12]Iwate Med. Univ., Morioka, Iwate, Japan, [13]Kanagawa Children's Med. Ctr., Yokohama, Kanagawa, Japan, [14]Tokyo Metropolitan Childrenâ□□s Med. Ctr., Fuchu, Tokyo, Japan, [15]Saiseikai Yokohama-shi Tobu Hosp., Yokohama, Kanagawa, Japan, [16]Natl. Ctr. for Child Hlth.and Dev., Setagaya, Tokyo, Japan, [17]Kanagawa Children's Med Ctr, Yokohama, Japan, [18]Natl. Ctr. for Child Hlth.and Dev.Keio Univ., Setagaya, Tokyo, Japan, [19]Osaka Women's and Children's Hosp., Osaka, Japan

Disclosure Block: **H. Suzuki:** None.

Among critically ill newborn infants who are admitted to neonatal intensive care unit, 5-10% have underlying genetic etiologies. Recently, genomic sequencing is increasingly recognized as a powerful diagnostic tool in critical care settings. We performed genomic analysis in critically ill neonates and infants to delineate the diagnostic efficacy of genomic sequencing by primary symptoms that prompted genomic testing and impact on the clinical management.

The study was conducted for 2 years between April 2019 and March 2021. A total of 85 patients were enrolled in this study. The final molecular diagnosis was provided in 41 of 85 subjects (48%). Four subjects had structural variations that were detectable by whole genome analysis. Based on the primary symptoms, metabolic phenotypes yielded the highest diagnostic rate of 67% (4/6 subjects), followed by renal (60%, 3/5 subjects) and neurological phenotypes (58%, 14/24 subjects). In the 41 subjects who were given final molecular diagnosis, 21/41 (51%) had change in the clinical management. New treatment options were identified in 6 subjects. Invasive testing (e.g., muscle biopsy and skin biopsy) were cancelled in 6 patients. Two subjects had critical changes in treatment plan; one had cancellation of organ transplantation and the other started preparation for organ transplantation.

Genomic analysis in critically ill neonates and infants showed high diagnostic rates for metabolic, renal, and neurological phenotypes. At least half of such infants would receive more optimal clinical care if the molecular diagnosis is made. For those who have neurometabolic phenotypes, genomic sequencing is recommended as the first tier of testing because of its high diagnostic yield and positive impact on clinical management.

PrgmNr 2302 - Residual risk for clinically significant copy number variants in pregnancies with normal NIPS

View session detail

Author Block: **L. Sagi-Dain**[1], I. Maya[2], D. Brabbing-Goldstein[2], L. Basel-Salmon[2], L. Salzer[2]; [1]Carmel Med. Ctr., Haifa, Israel, [2]Rabin Med. Ctr., Petach Tikva, Israel
Disclosure Block: **L. Sagi-Dain:** None.
Background: Chromosomal microarray analysis (CMA) is the recommended first-tier in pregnancies with sonographic anomalies, while in low-risk pregnancies this test detects clinically significant copy number variants (CNVs) in about 1%. As the constantly growing wide use of non-invasive prenatal screening (NIPS) facilitates the detection of chromosomal aberrations, defining the residual risk for abnormal CMA following normal NIPS is of importance for informed decisions regarding prenatal testing and screening options. The objective of our study was to shed light on this issue. **Methods:** CMA results of all pregnancies undergoing amniocentesis between the years 2013-2021 in large hospital-based laboratory were collected. Pregnancies with major sonographic anomalies or multiple fetuses were excluded. Clinically significant (pathogenic and likely pathogenic) CNVs were divided into: 3-NIPS-detectable (trisomies 13, 18 and 21), 5-NIPS-detectable (including sex chromosome aberrations), 5-NIPS and common microdeletion-detectable (including 1p36.3-1p36.2, 4p16.3-4p16.2, 5p15.3-5p15.1, 15q11.2-15q13.1, and 22q11.2 deletions), and genome-wide NIPS-detectable (including variants >7Mb). The theoretical residual risk for clinically significant CNVs was calculated following exclusion of NIPS-detectable findings. **Results:** Of the 8,099 pregnancies, clinically significant CNVs were demonstrated in 70 pregnancies (1.4%). The residual risk following theoretically normal NIPS was 1.2% (1/85) for 3-NIPS, 0.92% (1/109) for 5-NIPS, 0.88% (1/113) for 5-NIPS including common microdeletions, and 0.82% (1/122) for genome wide NIPS. In the subgroup of 4,048 pregnancies with advanced maternal age, the residual risk for clinically significant CNVs following theoretically normal NIPS ranged from 1.3% (1/75) for 3-NIPS to 0.8% (1/122) for genome wide NIPS. In 3,187 pregnancies of women younger than 35 years, this residual risk ranged from 0.7% (1/145) for 3-NIPS to 0.5% (1/198) for genome wide NIPS. The residual risk was highest for the 559 pregnancies with abnormal serum screening and 305 pregnancies with soft markers - about 2.2% (1/45) for 3-NIPS and 2.0% (1/50) for genome wide NIPS. **Conclusions:** The residual risk of clinically significant CNVs in pregnancies without structural sonographic anomalies is appreciable, and depends on NIPS extent, maternal age, the results of biochemical screening and presence of soft markers. This knowledge is important for the patients, the obstetricians and the genetic counselors, in order to facilitate informed decisions regarding prenatal testing and screening options.

PrgmNr 2303 - The potential of RNAsequencing for fetal malformation syndromes

View session detail

Author Block: **S. Aggarwal**[1,2], A. BHATTACHERJEE[3], I. Rungsung[2], V. S. Vineeth[2], V. S. Venkatapuram[1], A. Dalal[2]; [1]Nizam's Inst. of Med. Sci., Hyderabad, India, [2]Ctr. for DNA Fingerprinting and Diagnostics, Hyderabad, India, [3]The Ctr. for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, India

Disclosure Block: **S. Aggarwal:** None.

Objective: To assess the performance of RNA sequencing from cultured amniocytes as an investigative approach to study fetal malformation syndromes. Methods: Four fetuses with prenatal presentation of a fetal malformation syndrome underwent amniocentesis followed by establishment of amniocyte culture. Total RNA was extracted and converted into cDNA libraries further subjected to whole transcriptome sequencing. Paired end RNA-Seq data with 25X coverage was generated using Illumina HiSeq2500 sequencing platform. Good quality reads were processed to aligned to the human reference genome (GRCh37/hg19) using STAR RNA-Seq aligner (version 2.4.0.1). GATK was used to process the bam files to generate the variant call files and these were annotated using ANNOVAR, followed by standard variant filtering protocols. For quantification of transcripts and identification of alternative transcripts, data was analysed using Tuxedo pipeline. In one case, Gfold tool was used to estimate the differential gene expression with respect to a control sample. Standard log2foldchange cut-off was applied to screen the significantly differentially expressed genes in the data set, which was then used for various functional enrichment analysis.Results: The causative pathogenic variants were identified in all four cases from the RNAseq data. These were a homozygous nonsense variant in *NEK1* in a fetus with Short rib thoracic dysplasia type 6, a homozygous nonsense variant in *PLOD2* in a fetus with Bruck syndrome type 2, a homozygous missense variant in *ESCO2* in a fetus with Roberts syndrome and a de-novo heterozygous variant in *COL1A1* in a fetus with Osteogenesis imperfecta type II. The transcript analysis of the respective genes revealed nonsense mediated decay and alternative splicing events in the cases with *NEK1* and *PLOD2* variants, and no transcript abnormality with missense *ESCO2* and *COL1A1* variants. In the case with *NEK1* variant differential gene expression analysis as compared to control revealed insights into the possible downstream effects of the variant on overall transcriptome. Conclusions: RNAsequencing from cultured amniocytes is feasible, contains transcripts from multiple fetal tissues, and can be used to interrogate Mendelian phenotypes. This approach has the potential to identify the underlying putative variant, study the pathogenic potential of the variant as well as the transcriptome wide effect of the variant, thereby providing a window to study pathophysiology of various developmental Mendelian phenotypes.

PrgmNr 2304 - 3DeepMHC : 3D structure based Deep learning model for MHC-peptide binding prediction

View session detail

Author Block: **S. Jang**[1], B. Han*[2]; [1]Seoul Natl. Univ. Coll. of Med., SEOUL, Korea, Republic of, [2]Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of

Disclosure Block: **S. Jang:** None.

Cancer immunotherapy is a new method of treatment that helps the immune system recognize and attack the cancer cells with its own natural defense system. In fact, very few patients respond to this treatment. For immunotherapy to work, the patientâ□□s Major Histocompatibility Complex (MHC) should have the characteristics of combining with neoantigen of cancer. Lately, there have been many deep learning approaches to predicting MHC peptide binding affinity. However, previous methods only take into account the sequences or physicochemical properties, which suggests that they do not reflect the 3D genomic features. We developed the model called 3DeepMHC, which uses the information from the 3D structure of protein binding. Then we compared this 3DeepMHC with traditional deep learning-based methods. Our prediction model suggests the potential of clinical benefit for personalized neoantigen-based cancer immunotherapies more accurately.

PrgmNr 2305 - A 100,000 Genome Project haplotype reference panel of 156,390 haplotypes and the improved imputation of UK Biobank

View session detail

Author Block: **S. Shi**[1], S. Rubinacci[2], S. Hu[3], L. Moutsianas[4], A. Stuckey[4], C. Cabrera[5,6], V. Cipriani[5,6], D. P. Smedley[5,6], M. J. Caulfield[7], S. R. Myers[1], J. L. Marchini[8]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Univ. of Lausanne, Ecublens, Switzerland, [3]Novo Nordisk Res. Ctr., Oxford, United Kingdom, [4]Genomics England, London, United Kingdom, [5]William Harvey Res. Inst., London, United Kingdom, [6]Queen Mary Univ. of London, London, United Kingdom, [7]Genomics England, Queen Mary, London, United Kingdom, [8]Regeneron Genetics Ctr., Tarrytown, NY
Disclosure Block: **S. Shi:** None.
The Genomics England (GEL) 100,000 genome project has sequenced over 85,000 genomes across England. By using high coverage whole-genome sequencing (WGS), this constitutes the largest human genetic variation resource ever collected in the UK, and represents a near-complete characterization of genetic variation in the population. We generated a GEL haplotype reference panel, comprising 341 million autosomal variants and 156,390 haplotypes from diverse ancestries. We exploit both the sample size and relatedness structure among individuals, 61.3% of whom possess at least one sequenced first-degree relative, to allow high-precision haplotypic phasing.
We used 1000 Genomes WGS data to assess the imputation performance across ancestries, and observe improvements in some populations. In samples of British origin the mean imputation r2 at 0.01% allele frequency is 0.45, 0.67 and 0.74 when using the HRC, TOPMed and GEL reference panel. In samples of South Asian origin the mean imputation r2 at 0.01% allele frequency is 0.04, 0.24 and 0.61 when using the HRC, TOPMed and GEL reference panel.
We used the GEL reference panel to impute the UK Biobank dataset, that was previously imputed at 39 million autosomal variants, using an HRC+UK10K reference panel. It results in a ~6 fold increase in the number of imputed variants. Mean information scores at imputed SNPs, from the GEL and HRC-UK10K reference panels, were 0.65 and 0.61 respectively. At low allele frequencies the differences were larger. For example, for SNPs with allele frequency between 0.01% to 0.1% mean information scores were 0.88 and 0.66 for the GEL and HRC-UK10K reference panels respectively. This translates into an appreciable boost in power to detect associations. The GEL-imputed UK Biobank dataset is being made available to all approved researchers of the UK Biobank.
We will also report results of experiments of examine the implications for fine mapping and burden association tests in the context of imputed GWAS for blood pressure and other traits.

PrgmNr 2306 - A weighted selection probability to locate rare variants associated with highly correlated multiple phenotypes

View session detail

Author Block: **X. Liang**, H. Sun; Dept. of Statistics, Pusan Natl. Univ., Busan, Korea, Republic of
Disclosure Block: **X. Liang:** None.
In the past few decades, many statistical methods have been developed to identify rare variants associated with a complex trait or a disease. Recently, rare variant association studies with multiple phenotypes have drawn a lot of attentions because association signals can be boosted when rare variants truly associated with more than one phenotype. Most of existing statistical methods to identify rare variants associated with multiple phenotypes are based on a group test, where a gene or a genetic region is tested one at a time. However, these methods are not designed to locate individual rare variants within a gene or a genetic region. In this article, we propose a weighted selection probability to locate rare variants associated with highly correlated multiple phenotypes. Once a group-based test identifies the significant association of a group, the proposed method then computes selection probabilities of individual rare variants within the group to prioritize the rare variants. Selection probability represents selection frequency of nonzero regression coefficient in a regularization model using bootstrap sampling. Since the strength of an association vary in phenotypes for susceptible rare variants, selection probability weights can be computed based on a distribution of selection frequency of variants each phenotype. In our simulation study, we demonstrated that the weighted selection probability method outperforms unweighted selection methods in terms of true positive selection, when phenotype outcomes are highly correlated with each other. We also applied the proposed method to our genomic data set consisting of 10,783 rare variants and 13 correlated phenotypes. We identified some potentially susceptible rare variants which were missed by an association test with a single phenotype.

PrgmNr 2307 - Accurate imputation of human leukocyte antigens with CookHLA

View session detail

Author Block: **W. Choi**; Seoul Natl. Univ., Seoul, Korea, Republic of
Disclosure Block: **W. Choi:** None.
The recent development of imputation methods enabled the prediction of human leukocyte antigen (HLA) alleles from intergenic SNP data, allowing studies to fine-map HLA for immune phenotypes. Here we report an accurate HLA imputation method, CookHLA, which has superior imputation accuracy compared to previous methods. CookHLA differs from other approaches in that it locally embeds prediction markers into highly polymorphic exons to account for exonic variability, and in that it adaptively learns the genetic map within MHC from the data to facilitate imputation. Our benchmarking with real datasets shows that our method achieves high imputation accuracy in a wide range of scenarios, including situations where the reference panel is small or ethnically unmatched.

PrgmNr 2308 - An explainable machine learning method for polygenic scores using whole UK biobank genotype dataset

View session detail

Author Block: **R. Ota**, S. Morishita; The Univ. of Tokyo, Tokyo, Japan
Disclosure Block: **R. Ota:** None.

For studying polygenic diseases, GWAS has been largely contributing to identifying associated single nucleotide variants (SNVs); however, SNVs for polygenic diseases often have low odds ratio (1.1-1.3) and only explain a small proportion of heritability, which is called 'missing heritability.' To solve this problem, a number of statistical methods to estimate polygenic scores have been proposed. Khera et al. demonstrated that LDpred can identify a part of the population with a large odds ratio. However, these statistical methods typically use more than 10 thousand SNVs in P+T or more than 7 million SNVs in LDpred, making it hard to interpret the result to understand the mechanism of the disease. Several machine learning methods have been designed to select a limited number of relevant variables, such as SNVs, to output predictions that are easy to interpret while retaining high accuracy, and are therefore expected to be useful for the analysis of polygenic diseases. Indeed, there have been attempts to employ machine learning-based methods for risk prediction (Xu et al., 2020), though they output uninformative predictors that are either hard to interpret, almost as accurate as traditional statistical methods, or incapable of handling all SNVs in UK Biobank dataset due to large computational time. One promising machine learning method is Adaboost that adopts a linear model of a limited number of informative SNVs for better interpretation of the prediction model and is known to be infrequent to be overfitting. However, we found that Adaboost failed to capture low-frequent SNVs.

We developed an improved boosting method, Genoboost to capture low-frequent SNVs. Genoboost demands no information on LD beforehand but can implicitly exclude the effect of LD, which is the unique nature of boosting algorithm. We also developed a program that can process very large GWAS data such as the entire UK Biobank data, which contains about 300 thousand samples each of which has more than 7 million SNVs, in a reasonable amount of time (2.5 days).

For the Type 2 Diabetes data, letting methods use 50 SNVs, Genoboost identified 2.0% and 0.7% of the population whose respective odds ratios are 2 and 2.5 while P+T prediction did 1.2% and none and LDpred and Lassosum did less. The result demonstrates the merit of Genoboost for capturing a smaller number of SNVs relevant to the disease that applies to the wider population.

In conclusion, we propose the first machine learning method that can fully exploit the whole UK Biobank genotype dataset as far as we know, and can output better predictions than the state-of-the-art statistical methods, LDpred, P+T and Lassosum in terms of odds ratios and applicable population.

PrgmNr 2309 - Association tests of gene-environment interactions for time-to-event traits in family studies

View session detail

Author Block: **Y-F. Chiu**[1], C-Y. Lee[2]; [1]Natl. Hlth.Res. Inst.s, Zhunan, Taiwan, [2]Natl. Hlth Res Inst, Zhunan, Miaoli, Taiwan
Disclosure Block: **Y. Chiu:** None.
Complex disease often involves both genetic and environmental factors. Investigating gene by environment (G-E) interactions is therefore an essential step to dissect disease etiology. Analysis of gene-environment (G-E) interactions for rare variants for survival traits is challenging due to lack of statistical power. Family-based study designs help improve statistical power as functional variants are likely to be aggregated in family studies enriched with affected member. Previously, we developed novel pedigree-based burden and kernel association tests for time-to-event outcomes with right censoring for pedigree data, referred to FamRATS (**Fam**ily-based **R**are variant **A**ssociation **T**ests for **S**urvival traits). In the present study, we extended our developed tests for survival outcomes to jointly test for genetic main effects and G-E interactions between a set of rare variants and environmental factors. Cox proportional hazard models were employed to relate a time-to-event trait with interactions between rare variants and environmental factors. The proposed tests do not require for estimating individual effects of rare variants and their G-E interactions. Adjustments were made for the fixed effects of confounding factors. Comprehensive simulation studies were conducted to evaluate the proposed tests in terms of type I error rates and statistical power. Application to a real data example for time-to-event traits was used for an illustration.

PrgmNr 2311 - Correction for sample overlap, winner's curse and weak instruments bias in two-sample Mendelian Randomization

View session detail

Author Block: **N. Mounier**[1,2], Z. Kutalik[1,2,3]; [1]Univ. Ctr. for Primary Care and Publ. Hlth., Lausanne, Switzerland, [2]Swiss Inst. of Bioinformatics, Lausanne, Switzerland, [3]Dept. of Computational Biology, University of Lausanne, Switzerland
Disclosure Block: **N. Mounier:** None.

Introduction:
Inverse-variance weighting (IVW) two-sample Mendelian Randomization (MR) is the most widely used method to estimate the causal effect of an exposure on an outcome. However, the resulting causal effect estimates may suffer from different biases due to sample overlap, winnerâ□□s curse and weak instruments.

Methods:
Assuming a spike-and-slab genomic architecture, we analytically derived the bias of such estimate, which can be quantified using only summary statistics. Hence, we propose a correction of the IVW-MR estimate and compared it against its uncorrected counterpart under a wide range of simulations settings. Finally, we performed IVW-MR based on summary statistics for body mass index (BMI) and systolic blood pressure (SBP) obtained from overlapping samples(N_{BMI}=686,128, N_{SBP}=340,159, $N_{overlap}$= 340,159) and corrected the obtained causal effect estimates using our method.

Results:
Using simulated data, we observed that when the confounder and the causal effect are acting in the same direction, IVW-MR effects are overestimated for fully-overlapping samples and underestimated for non-overlapping samples. When they are acting in opposite directions, observed effects are underestimated for all overlaps because the three sources of biases are towards the null. In all the explored scenarios, our correction reduced bias (up to 30 folds). Using summary statistics for real data, our method revealed that IVW-MR causal effects of BMI on SBP and of SBP on BMI were both significantly overestimated (by 15% and 10% respectively).

Conclusions:
We developed a method to correct causal effect estimates for sample overlap, weak instrument bias and winnerâ□□s curse simultaneously using only summary statistics.

PrgmNr 2313 - Genome-wide classification of epigenetic signal reveals regions of enriched heritability in complex immune traits

View session detail

Author Block: **M. Stricker**[1], W-Y. Cheng[2], S. Nahkuri[3], P. Palamara[1]; [1]Univ. of Oxford, Oxford, United Kingdom, [2]Roche Pharma Res. & Early Dev. Informatics, Roche Innovation Ctr., New York, NY, [3]Roche Pharma Res. & Early Dev. Informatics, Roche Innovation Ctr., Zurich, Switzerland
Disclosure Block: **M. Stricker:** Other; Hoffmann-La Roche.
Epigenetics is known to play a key role in the regulation of adaptive and innate immune-system (AIIS) relevant genes that dictate the therapeutic course for many diseases such as cancer and COVID-19. We built a model that leverages epigenetic data to perform supervised gene classification, with the goal of detecting AIIS genes. We leveraged known similarities between epigenetic states to encode high-dimensional epigenetic data (Epigenome Roadmap, Kundaje et al 2015; ChromHMM, Ernst & Kellis 2012) for different human tissues into images. We then trained a convolutional neural network from a handcrafted list of 477 known AIIS genes to perform image recognition, achieving a testing accuracy of 0.93 (SE 0.03). We used the trained model to scan the human genome for new AIIS regions and detected 1964 putative loci, of which 1129 do not have an associated Gene Ontology (GO) term. Some of these predicted novel AIIS regions harbour immune-cell specific alternative transcription variants, suggesting underlying biological mechanisms.
To evaluate the role played by these predicted AIIS regions in the genetic architecture of complex immune-related traits, we built a genome-wide annotation that reflects our modelâ□□s confidence of AIIS relevance at each locus. Regions with high predicted AIIS relevance (score > 0.5) were found to be enriched for GO terms related to immune-related function (p = 7e-21). We used linkage disequilibrium score regression (LDSC) coupled with genome-wide association summary statistics for 176 traits (average N = 262k) to test whether our AIIS annotation is predictive of regional heritability for these traits. We detected a significant heritability enrichment (LDSC |Ï□*| p To confirm the specificity of the detected enrichments for immune-related phenotypes, we performed a meta-analysis of 64 independent traits, including 4 autoimmune diseases. For all autoimmune diseases, the heritability enrichment remained strong (|Ï□*| > 0.5; p These results underscore the promise of leveraging machine learning algorithms and large epigenetic datasets to detect genomic regions implicated in specific classes of heritable traits and diseases.

PrgmNr 2314 - Integrating facial and genomic data for subtyping individuals in the presence of confounders via individual-specific network analysis

View session detail

Author Block: **Z. Li**[1], D. Duroux[2], M. D. Shriver[3], S. Walsh[4], S. M. Weinberg[5], P. Claes[6], B. MÃ¼ller-Myhsok[7], K. Van Steen[1,2]; [1]BIO3 â Lab. for Systems Med., Katholieke Univ.it Leuven, Leuven, Belgium, [2]BIO3 â Lab. for Systems Genetics, GIGA-R Med. Genomics, Univ. of LiÃ¨ge, LiÃ¨ge, Belgium, [3]Penn State Univ, University Park, PA, [4]Indiana Univ. Purdue Univ. Indianapolis, Indianapolis, IN, [5]Univ of Pittsburgh, Pittsburgh, PA, [6]Lab. for Imaging Genetics, Katholieke Univ.it Leuven, Leuven, Belgium, [7]Group of Statistical Genetics, Max Planck Inst. of Psychiatry, Munich, Germany
Disclosure Block: **Z. Li:** None.

With the collection of vast amounts of heterogeneous data, an increasing number of individuals can be featured by multiple data types, giving richer and complementary information than single data views. This is particularly useful in the context of precision medicine in which the individual plays a central role and benefits can be gained from grouping individuals by âsimilarityâ. Such similarity assessments in sample collections is typically part of latent class analyses, in particular multi-view clustering efforts. The issue of confounded clustering is dealt with a posteriori, if at all. In more recent efforts, individual-specific networks (here, referring to networks with subject-specific edge weights) are constructed and processed, usually focussing on a single data source. In this work, we propose a novel analysis workflow for data integrative subtyping of individuals from two data modalities, in the presence of confounders. Clustering is performed at two possible levels: 1) multi-view components that are extracted from each data view by (sparse) Canonical Correlation Analysis (CCA), and 2) individual-specific (sub-)networks (ISNs) that are derived from the feature similarity matrix following CCA on the entire dataset. ISN nodes (features) represent both data views; edges represent perturbations of an individual to the entire sample network. Special attention is given to optimize the selection of subnetworks when clustering (sub-)ISNs; the latter via hierarchical arguments and similarity measures between graphs (e.g. shortest path kernel distance). The number of subgroups is derived via a recursive strategy, relying on an adaption of distance-based ANOVA and permutation testing. In the presence of known confounders, the workflow consists of 1) quantifying the impact on subtyping (e.g. comparing final clusterings when CCA was informed versus uninformed by confounders), and 2) adjusting for the undesired effects (e.g. via Partial Least Squares regression directly on the original input data, via Kernel Conditional Clustering at the clustering stage itself).For illustrative purposes, we apply the outlined workflow on real-life data from a US cohort of 4680 individuals, assuming age as confounder: facial images are hierarchically grouped into 63 segments and represented by principal components; genetic markers are mapped to genes by genomic positions. All obtained clusterings are compared (f.i. Variation of Information metric). Our work shows the importance of eliminating undesired factors driving individual subtyping and offers unprecedented approaches via multi-view individual-specific networks.

PrgmNr 2315 - Mendelian Randomization with repeated measurement using Taiwan Biobank

View session detail

Author Block: **J-J. Sie**[1], C. S. Fann[2], L. Ie-Bin[3]; [1]Dept. of Mathematics, Natl. Changhua Univ. of Ed., Changhua County, Taiwan, [2]Inst Biomed Sci, Acad Sinica, Taipei, Taiwan, [3]Graduate Inst. of Statistics and Information Sci., Natl. Changhua Univ. of Ed., Changhua County, Taiwan
Disclosure Block: **J. Sie:** None.
Mendelian randomization (MR) controls for the confounding and reversed causality problem in observational studies. Genetic variants as instrumental variables (IV) are utilized to mimic the randomization of exposure. Many modified MR methods have been developed to cope with a variety of study designs and scenarios, e.g., multivariable MR methods, MR methods controlling for pleiotropy, etc. As common practice is to use the mean as a representation for repeated measures, to our knowledge, no MR as of yet has been developed to effectively manage repeated measurements or longitudinal data. In this study, we propose an MR method incorporating Generalized Estimation Equation (GEE) for data with repeated measurements, and compare its performance with conventional MR approaches. We applied this method using data from Taiwan Biobank, which contains more than 120,000 genotyped individuals (March 2021). Among them, more than 22,000 have one follow-up measurements. The range of follow up is between 3 to 6 years (mean = 3.84); age between 30-70 years old with about 64% female. In order to prove our concept, the well known causality between BMI and T2D was chosen. For the baseline data, the mean and standard deviation of BMI were (24.1, 3.59) and the T2D prevalence was 5.13%. For the follow up data, they were (24.3, 3.72) for BMI mean and standard deviation and T2D prevalence was 7.5%. If mean was used as a proxy for two measurements of BMI, OR from MR was 1.549 with se = 0.161 and p = 0.0066. The obtained IVs from this model were used for the subsequent analyses. When using the baseline BMI, OR = 1.316 with se = 0.135 and p = 0.0413; OR = 1.575 with se = 0.152 and p= 0. 0027 when follow up BMI was used. By using MR-GEE model, OR = 1.52 with se = 0.132 and p= 0. 00146.In comparison of these MR results from different models, the causality between BMI and T2D was all significant. The estimated OR were very similar; however, standard error and p value from MR-GEE were smallest. MR-GEE utilizes the longitudinal measurements of both exposure and disease status by taking into account of their correlation, which may make it a more powerful approach than using mean as a proxy, in terms of producing smaller standard error. The causality between other phenotypes are being tested using our proposed MR-GEE model. The limitation of Taiwan Biobank data is that there is only one follow up data available. Simulated follow up data will be used in future studies.

PrgmNr 2316 - Significance in individual-specific networks: assessment, methods and characteristics

View session detail

Author Block: **F. Melograna**[1], D. Duroux[2], N. van Best[3,4], M. Mommers[5], J. Penders[3,6], F. Stella[7], K. Van Steen[1,2]; [1]BIO3 â Lab. for Systems Med., Katholieke Univ.it Leuven, Leuven, Belgium, [2]BIO3 â Lab. for Systems Genetics, GIGA-R Med. Genomics, Univ. of LiÃ¨ge, LiÃ¨ge, Belgium, [3]Sch. of Nutrition and Translational Res. in Metabolism (NUTRIM), Dept. of Med. Microbiol., Maastricht Univ., Maastricht, Netherlands, [4]Inst. of Med. Microbiol., RWTH Univ. Hosp. Aachen, RWTH Univ., Aachen, Germany, [5]Dept. of Epidemiology, Care and Publ. Hlth.Res. Inst. (CAPHRI), Maastricht Univ., Maastricht, Netherlands, [6]Care and Publ. Hlth.Res. Inst. (CAPHRI), Dept. of Med. Microbiol., Maastricht Universit, Maastricht, Netherlands, [7]Dept. of Informatics, Systems and Communication, Univ. of Milan-Bicocca, Milan, Italy

Disclosure Block: **F. Melograna:** None.

Individual-specific networks (ISNs) are graphical structures for which at least one element type (node, edge) has individual-specific labels or weights. In this work, we assume ISNs to be undirected networks with subject-specific edge weights and node values, with edges as a major information source (as in Kuijjer et al., 2019). ISNs are valuable tools to understand individual-to-individual heterogeneity, a key component in Precision Medicine. One of the shortcomings of edge-oriented ISNs, hampering valorization, is that these are usually not encapsulated in a framework that assesses whether it is worthwhile to continue working with an individual-specific network. In other words, it is of interest to know when an ISN is â significantâ. The notion of significance depends on the context but often reduces to how significantly different a leave-one-out-individual network is from the full network (i.e. no individuals left out). Most approaches deal with the significance of an individualâs perturbation, edge by edge, often not considering proper standard errors for test statistics that contrast two strongly dependent samples. Moreover, it defeats the effort of deriving a graphical structure for each individual that can later serve as input to predictive or descriptive models. In an extensive simulation study, we investigate the impact of distributional assumptions of edge weights in a full network from which ISNs are derived, on ISN significance. Deviations from normality are emphasized to cover networks built via more general association than âcorrelationsâ. We explore multiple ISN significance assessment methods, including a newly proposed method ISN-LOO, that involves frequent resampling from a multivariate normal distribution, with covariance structure templated on the observed full network. We consider both edge or module driven significance assessment, relying on appropriate distance measures between an edge (respectively, a module) via subtraction of edge weights for the full and LOO model (respectively, via graph kernels). The novel approach is compared to outlier detection approaches (e.g., proximity-based algorithms and influence analyses). Testing for multiple edges (respectively multiple modules) is controlled at 5%. Our results show the added value of ISN-LOO when deviations from normally distributed full edge weights are limited. Also, modular significance assessment is most powerful for more complex network structures. We furthermore use real-life data from 16S microbiome profiles of 69 newborns from the Dutch LucKi cohort, and illustrate the utility of ISN significance testing to highlight individuals of interest.

PrgmNr 2317 - SNPs tagging structural variations help to reveal widespread role of CNVs in complex traits and gene expression

View session detail

Author Block: **M. Lepamets**[1], K. Alasoo[2], M. Kals[1,3], U. VÃµsa[1], R. MÃ¤gi[1]; [1]Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [2]Univ. of Tartu, Tartu, Estonia, [3]Inst. for Molecular Med. Finland, FIMM, HiLIFE, Univ. of Helsinki, Helsinki, Finland
Disclosure Block: **M. Lepamets:** None.
Genome-wide association studies (GWAS) have been successful at identifying links between single nucleotide polymorphisms (SNPs) and complex traits. Pinpointing true causal markers, however, has been challenging, and can be even more so in loci where SNP effects reflect the underlying structural variation. In order to recognise such loci, we identify copy number variable (CNV) regions that are tagged by SNPs (tagSNPs) and evaluate the abundance of those tagSNPs among known GWAS and expression quantitative trait loci (eQTLs).
We detected CNVs using Genome STRiP pipeline from ~2,300 Estonian Biobank (EstBB) whole-genome sequencing samples. After initial quality control, ~9M SNPs (MAF>0.01) and 3,631 common CNV regions were included in the tagSNP identification process. We discovered that 52.2% (N=1,895) of CNVs had tagSNPs (R^2>0.8) in their 1Mb radius.
First, we overlapped the tagSNPs with known associations from GWAS Catalog and found that for 285 CNVs the tagSNPs were in high LD (R^2>0.8) with a GWAS top hit. These make up about 1.6% of all GWAS Catalog associations that passed our filtering criteria, 3.6 times more than expected by chance (P-300). The most frequent associations were with BMI (18 CNVs) and different cognitive ability traits (in total 24 CNVs). Second, we analysed the overlap of tagSNPs and 95% credible sets (CS) for eQTLs from eQTL Catalogue. We discovered 309 unique CNV regions for which the tagSNPs covered over 90% of a single CS in at least one dataset/tissue. In 122 (39.5%) cases the CNV physically overlapped with the corresponding gene. Furthermore, we conducted genetic fine-mapping analysis on ~500 EstBB samples with: a) SNPs only and b) SNPs and CNVs combined. We found that in 28.2% (N=50) of CS that originally contained tagSNPs (N=177), the CNVs were chosen over or in addition to the tagSNPs in the second run as potentially causal.
Altogether, we found that 471 (24.9%) CNV regions which were tagged by SNPs showed evidence of being an underlying variant behind either GWAS or eQTL SNP associations. Considering the pronounced impact on gene function and dosage that CNVs can have, they are strong candidates for causal variants. Therefore, tagSNP information is essential when designing functional follow-up experiments based on SNP results.

PrgmNr 2318 - Statistical selection method to identify pleiotropic variants associated with both quantitative and qualitative traits

View session detail

Author Block: **K. Kim**, H. Sun; Pusan Natl. Univ., Busan, Korea, Republic of
Disclosure Block: **K. Kim:** None.

In recent genetic association studies, statistical methods to identify pleiotropic variants associated with multiple phenotypic traits have been developed, since susceptible variants with small or moderate effects are rarely detected by association methods based on a single trait. However, most of the existing methods to identify pleiotropic variants are designed for only quantitative traits even though pleiotropic variants are often associated with both quantitative and qualitative traits. This is a statistically challenging problem because there does not exist an appropriate multivariate distribution to model both quantitative and qualitative data. There are some meta-analysis methods which basically integrate summary statistics of individual variants associated with either a quantitative or qualitative trait. However, these methods cannot account for correlations between genetic variants. In this article, we propose new selection method to prioritize individual variants associated with both quantitative and qualitative traits. For individual traits, regression coefficients of elastic-net regularization are first estimated and then they are additively combined to compute selection probability of individual variants. In our extensive simulation studies where either homogeneous or heterogeneous variant effects on both quantitative and qualitative traits were considered, we demonstrated that the proposed method outperforms the existing meta-analysis methods in terms of true positive selection. We also applied the proposed method to real genomic data with both quantitative and qualitative traits.

PrgmNr 2319 - The Alpha variant: A comparison for risk for hospitalization and case-mortality rate for younger and older patients

View session detail

Author Block: **S. Cetin**[1,2], A. Ulgen[3]; [1]Ã¼nv, Tokat, Turkey, [2]Tokat GaziosmanPasa Univ., Biostatistics, Faculty of Med., Tokat, Turkey, [3]Faculty of Med., Girne American Univ., Karaman, Cyprus
Disclosure Block: **S. Cetin:** None.

Knowing the risks for hospitalization and case-fatality-rate for emerging SARS-CoV-2 variants for different age groups is necessary for hospital management and vaccination planning. It is therefore important to collect and analyze health data related to the variants. We have obtained more than 3700 COVID-19 patients (3100 outpatients and 600 inpatients) where about 30% are infected by Alpha variant. Both logistic regression and cause-specific Cox survival analysis of competing-risk is run on inpatients to examine the impact of the Alpha variant on hospitalization and on mortality, conditional on other factors. Descriptive statistics is used to characterize different subgroups. We observed that the Alpha variant is over-represented in inpatients than outpatients so carrying the Alpha variant increases the chance for hospitalization. The impact of the Alpha variant on mortality seems to depend on the patientâ□□s age. For age *70 group, the case-fatality-rate is 0.84% (5.3%) for patients without (with) the Alpha variant (Fisherâ□□s test P-value = 2.4 Ã□10â□□10). For age â□¥ 70 group, the trend is opposite: the case-fatality-rate is 31.5% (13.6%) for patients without (with) Alpha variant (Fisherâ□□s test P-value = 0.0016). The two opposite trends would cancel each other, making other analyses such as cause-specific Cox regression and logistic regression non-significant. The Alpha variant increases the risk for hospitalization, increases the case-fatality-rate for lower age group, and decreases the case-fatality-rate for the upper age group. It is therefore imperative to vaccinate young, middle-aged, and early senior population to counter the impact of wave of the Alpha variant.*

PrgmNr 2320 - Time to mortality and time to discharge from the hospital in Covid-19 data: A Turkish study

View session detail

Author Block: **A. Ulgen**; Dept. of Biostatistics, Faculty of Med., Girne American Univ., Karmi, Cyprus
Disclosure Block: **A. Ulgen:** None.
COVID-19 survival data presents a special situation where not only the time-to-event period is short, but also the two events or outcome types, death and release from hospital, are mutually exclusive, leading to two cause-specific hazard ratios ($csHR_d$ and $csHR_r$). The eventual mortality/release outcome is also analyzed by logistic regression to obtain odds-ratio (OR). We have the following three empirical observations: (1) The magnitude of OR value is an upper limit of the $csHR_d$ value: |log(OR) | â¥ | log($csHR_d$)|. This relationship between OR and HR could be understood from the definition of the two quantities in an approximation; (2) $csHR_d$ and $csHR_r$ point in opposite directions: log($csHR_d$) . log($csHR_r$) d and $csHR_r$: $csHR_d$ ~ 1/$csHR_r$. Though an approximate reciprocal trend between the two hazard ratios is an indication that the same factor causing faster death also lead to slow recovery by a similar mechanism, and vice versa, a quantitative relation between $csHR_d$ and $csHR_r$ in this context is not obvious. These observations will help future analyses of COVID-19 data, as well as other data of similar nature, in particular if the deceased samples are lacking. In particular, if the death rate is moving towards zero with future life-saving drugs, it is still possible to obtain $csHR_r$ using large number of surviving patients; then inferring $csHR_d$ from $csHR_r$.

PrgmNr 2321 - TransformerPRS: a deep learning-based polygenic risk model using bidirectional transformers derived from a language model

View session detail

Author Block: **J. Kim**[1], S-H. Jung[2,3,4], M. Kim[3,4], K. Nho[5], L. Shen[3,4], K-A. Sohn[1,6], H-H. Won[2,7], D. Kim[3,4], for the Alzheimerâ□□s Disease Neuroimaging Initiative; [1]Dept. of Computer Engineering, Ajou Univ., Suwon, Korea, Republic of, [2]Dept. of Digital Hlth., SAIHST, Sungkyunkwan Univ., Samsung Med. Ctr., Seoul, Korea, Republic of, [3]Dept. of Biostatistics, Epidemiology and Informatics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [4]Inst. for BioMed. Informatics, Univ. of Pennsylvania, Philadelphia, PA, [5]Dept. of Radiology and Imaging Sci., Indiana Univ. Sch. of Med., Indianapolis, IN, [6]Dept. of Artificial Intelligence, Ajou Univ., Suwon, Korea, Republic of, [7]Dept. of Hlth.Sci. and Technology, SAIHST, Sungkyunkwan Univ., Samsung Med. Ctr., Seoul, Korea, Republic of
Disclosure Block: **J. Kim:** None.

Polygenic risk score (PRS) is the cumulative, mathematical aggregation of risk derived from the contributions of many DNA variants across the genome. Recent studies have demonstrated the utility of PRS on identifying individuals who are at high risk of disease and could benefit from early preventative intervention. However, there are several limitations of PRS; 1) only capturing the marginal effect of variants, 2) only reflecting linear effects of variants, and 3) no consideration and interpretation of any interactions between genes. In this study, we propose *TransformerPRS*, a deep learning-based model using a transformer module derived from a language model, which combines the concepts of PRS and interactions between genes to overcome the limitations of PRS. In our approach, we first introduce a monogenic risk score (MRS) by aggregating risks of variants in a gene and transform the MRSs into a sequence of risky genes and a sequence of protective genes. *TransformerPRS* takes these as input and learns the optimal representation of genes to predict disease risk by incorporating the interaction between input tokens. We also propose *TransformerPRS with pre-trained gene embeddings* using biological knowledge provided by *HiG2Vec* which is based on gene embedding. For the experiments, we used whole genome-sequencing data from ADNI-WGS-2 with ADSP Follow-Up Study (n=1,566) and the AD GWAS summary statistics data from an independent IGAP cohort (with ADNI-ADSP samples removed). As a result, conventional PRS (with p-value threshold=1E-04) showed an area under the receiver operating characteristics curve (AUC) of 0.658 Â± 0.034 (mean Â± std) to classify patients with Alzheimerâ□□s disease and cognitively normal participants. Both *TransformerPRS* and *TransformerPRS with pre-trained gene embeddings* outperformed it by AUC of 0.680 Â± 0.033 and 0.691 Â± 0.034, respectively. In addition, the self-attention module in a transformer block identified important features and their interactions. Our models can improve disease risk prediction by providing information on which genes and interactions between genes have an important impact on disease prediction, which are missed by conventional PRS.

PrgmNr 2322 - Using a digital health tool to analyse UK primary care electronic health records to identify potential cases of undiagnosed Marfan syndrome

View session detail

Author Block: **P. Ravichandran**, O. Buendia, W. Evans, L. Menzies; Mendelian, London, United Kingdom
Disclosure Block: **P. Ravichandran:** None.
Background: Marfan syndrome (MS) is a genetic disorder that affects connective tissue. It is caused by mutations in the fibrillin-1 gene. Aortic dissection is a significant cause of premature mortality. Early identification enables monitoring, risk factor modification and timely surgery. We used a digital health case-finding tool, MendelScan, to identify potential cases of MS in UK primary care electronic health records (EHR). Method: Firstly, we identified MS as a suitable rare disease to digitalise using our internal disease scoring metrics. We digitised a criterion based on the 2010 revised Ghent nosology to create a digital algorithm, refining the algorithm using common coded clinical features identified in the EHR of patients with established diagnoses. Each identified case underwent a manual internal clinical review, assessing the full structured (no free text) information held in the EHR for each case and evaluating whether there was a strong enough suspicion of MS to recommend further investigation. Results: Using a sample of 501,188 EHR, our algorithm identified 39 potential cases (including 3 previously diagnosed cases); internal clinical review refined the number of potential previously undiagnosed cases to 10.There were 47 previously diagnosed cases within our sample, of which our digital health tool identified 3. Our algorithm flagged 2 of these cases in advance of their diagnosis being recorded in their EHR; by 1 year and 2 years. Conclusion: We have demonstrated that using a digital health case-finding tool, with an algorithm developed from the Ghent nosology, we can identify credible cases with possible undiagnosed MS for further investigation.

PrgmNr 2323 - A machine-learning-based prediction of disease endpoints from administrative health data accounts for misclassification and disease liability in GWAS

View session detail

Author Block: **L. Eick**[1], S. Jukarainen[1], M. Cordioli[1], T. T. Kiiskinen[1,2], FinnGen, A. Ganna[3,1]; [1]FIMM, Inst. for Molecular Medicin Finland, Helsinki, Finland, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Harvard Med. Sch. and Massachusetts Gen. Hosp., Boston, MA
Disclosure Block: **L. Eick:** None.
Genome wide association studies (GWAS) assume there is a correct classification of subjects into cases and controls. Nonetheless, misclassification is expected, especially in biobanks that rely on administrative health data. Additionally, the genetic liability to a disease might best be viewed as a continuous value rather than a binary one. In order to address potential misclassification and increase the statistical power of GWAS, we implement an XGBoost based classifier. Instead of binary disease labels, it outputs a continuous liability measure based on diagnoses and medication usage and takes the result to perform a GWAS.
We analyzed data from the FinnGen study (n=253,072) based on Finnish registries with a follow-up of up to 50 years. We used 4003 diseases and 445 different medication classes as predictors for the classifier. Our method can be used for any disease present in FinnGen, and we demonstrate the method using ischemic stroke. We performed a GWAS over all individuals of FinnGen and defined as cases those individuals that were at high stroke risk according to our machine-learning-based prediction. We identified three GW-significant loci that were not found in GWAS that used the original case control classification. All three loci are known to be associated with stroke based on results from the largest stroke GWAS meta-analysis to date.
Our method expand GWAS possibilities accounting for potential misclassification and leveraging disease liability in individuals that are not yet diagnosed.

PrgmNr 2324 - Characterization of the human *ABO* genotypes and their association to inflammatory and cardiovascular disease

View session detail

Author Block: **J. R. HÃ¶glund**, T. Karlsson, T. Johansson, W. E. Ek, Ã. Johansson; Uppsala Univ., Uppsala, Sweden
Disclosure Block: **J.R. HÃ¶glund:** None.
The *ABO* gene contains three major alleles that encodes different antigens; A, B, and O, which determine an individualâs blood group. Previous studies have primarily focused on identifying associations between ABO blood groups and diseases risk. Here, we sought to test for association between *ABO* genotypes (OO, OA, AA; OB, BB, and AB) and a large set of diseases and disease-related protein biomarkers. We conducted association studies in two European cohorts; UK Biobank and NSPHS. We first tested for association by conducting a likelihood ratio test, testing whether ABO contributed significantly to the risk for 24 diseases, and 438 plasma proteins. For phenotypes with FDR ABO genotypes using logistic or linear regression. Our study confirmed previous findings of a strong association between *ABO* and cardiovascular disease, and provide additional evidence of significant differences between heterozygous and homozygous allele carriers for pulmonary embolism, deep vein thrombosis, but also for von Willebrand factor levels. Additionally, we found that *ABO* contributed significantly to 39 plasma proteins, of which 24 have never been linked to the *ABO* locus before. These results show the need of incorporating *ABO* genotype information in the consultation and management of patients at risk, rather than classifying patients into blood groups. Furthermore, the results indicated an additive effect between genotypes.

PrgmNr 2325 - Effects of *GLP1R* coding variation on random glucose levels and functional responses to GLP-1R agonists using whole exome sequencing data

View session detail

Author Block: **J. G. Maina**[1,2], A. Ulrich[2], Z. Balkhiyarova[2,3], R-M. Rujan[4], G. Deganutti[4], C. A. Reynolds[4], M. Kaakinen[2], I. Prokopenko[1,2,3], B. Jones[5]; [1]UMR 8199 - EGID, Inst. Pasteur de Lille, CNRS, Univ. of Lille, Lille, France, [2]Dept. of Clinical and Experimental Med., Sch. of BioSci.s and Med., Univ. of Surrey, Guildford, United Kingdom, [3]Ufa Federal Res. Ctr. Russian Academy of Sci., Ufa, Russian Federation, [4]Ctr. for Sport, Exercise and Life Sci., Faculty of Hlth.and Life Sci., Coventry Univ., Coventry, United Kingdom, [5]Dept. of Metabolism, Digestion and Reproduction, Faculty of Med., Imperial Coll. London, London, United Kingdom
Disclosure Block: **J.G. Maina:** None.
Genetic imputation has been useful in enlarging the number of testable variants for genome-wide association studies (GWAS). However, imputation and genotyping accuracy for rare variants (minor allele frequency, MAF5%). The usefulness of both array genotyping and imputation in establishing the role of rare variants in complex diseases is therefore thought to be limited. In this study, we compared the estimated effects of glucagon-like peptide-1 receptor (*GLP1R*) gene missense single nucleotide polymorphisms (SNPs), derived from imputed/array genotype or whole exome sequencing (WES) UK Biobank data, on random glucose (RG). We performed functional and structural studies to explain these effects and identify how patients with type 2 diabetes might respond differently to GLP-1R agonist drugs. We used the latest WES data release from the UK Biobank to test for association between *GLP1R* coding variants and natural log transformed RG adjusted for age, sex and time since last meal under an additive model of genetic effects in 165,818 unrelated European individuals. Further, we ran the same RG GWAS model in the UK Biobank (n=401,810) using the Haplotype Reference Consortium (HRC) imputed panel. For selected coding missense *GLP1R* variants we experimentally assessed G protein coupling and endocytosis in response to endogenous and FDA-approved therapeutic GLP-1R agonists, and performed molecular dynamics simulations of the oxyntomodulin-bound human GLP-1R using a recently described cryo-EM structure. 853 *GLP1R* coding variants were detected in WES data while 18 variants were detected in the HRC-imputed GWAS data. RG effects estimated in the larger imputed dataset were directionally consistent with the WES dataset, but effect size consistency was highest for commoner variants (e.g., rs10305492; protein coding consequence A316T, MAF = 1.5%, $BETA_{WES}$ = -0.011, SE_{WES} = 0.001; $BETA_{Genotyped}$ = -0.011, $SE_{Genotyped}$ = 0.002). The RG-lowering A316T variant showed increased agonist-induced coupling to stimulatory G proteins and endocytosis, which was explained by a structural alteration in the nearby hydrogen bond network. Experimentally determined gain-of-function for A316T applied more to certain GLP-1R agonists (e.g., tirzepatide) than to others (e.g., exenatide). Sensitivity analysis using WES data confirmed imputed data for common variants on RG effect. Our observation of ligand-specific responses to *GLP1R* variants such as A316T provides a mechanism that can explain why some individuals respond better or worse to particular GLP-1R-targeting drugs and a possible framework for T2D treatment stratification.

PrgmNr 2326 - Genetic and environmental determinants of drug adherence and drug purchasing behaviour

View session detail

Author Block: **M. Cordioli**[1], S. Jukarainen[1], T. T. Kiiskinen[1], FinnGen Study, S. Ripatti[1], A. Ganna[2,1]; [1]Inst. for Molecular Med. Finland, Univ. of Helsinki, Helsinki, Finland, [2]Harvard Med. Sch. and Massachusetts Gen. Hosp., Boston, MA
Disclosure Block: **M. Cordioli:** None.
One of the major factors behind the efficacy of pharmacological treatments is patients adherence to the prescribed therapy regimen. While demographic and socioeconomic factors may play a role in determining adherence, there have been few investigations on the potential effects of genetic variation on drug adherence. Drug adherence is difficult to measure because, unlike prescriptions, drug purchases are often not recorded in a centralized database. The Finnish drug purchase registry offers a unique resource towards this aim, covering a time span of more than 20 years and including the dose and quantity of every reimbursable prescription medication purchased at a pharmacy.
By leveraging genetic data from the FinnGen study (N=309,154) and data from the Finnish drug purchase registry (63,084,035 total purchases), we provide a systematic investigation of adherence determinants across multiple diseases and medications.
For each drug class of interest, we reconstructed individual drug trajectories and defined adherence as the ratio between the total purchased quantity and the length of the trajectory. We then ran a GWAS of adherence to each medication.
Here we report the results for statins (2,821,628 purchases, N=90,617) and blood pressure (BP) medications (4,512,383 purchases, N=99,793). We replicated the analysis for statins using prescription data from the UK Biobank as a proxy for adherence (N=37,783) and meta-analyzed with FinnGen results. For both medications, we estimated the genetic correlation (r_g) between adherence and 27 publicly available traits.
Overall, the results suggest adherence pertains to general behavioural aspects rather than to biological factors. For example, higher adherence to statins resulted to be genetically associated with higher educational attainment (r_g=0.22, P=3.3x10^{-4}) a lower risk tolerance (r_g=-0.36, P=6.6x10^{-5}) and lower openness to experiences (r_g=-0.3, P=6.8x10^{-4}). Psychiatric traits such as schizophrenia, bipolar disorder and depression also showed a negative genetic correlation with adherence. Furthermore, in both UK Biobank and FinnGen, well-known variants in statins pharmacogenetics (*ABCG2, HMGCR, SLCO1B1*) showed no effect on adherence.
By extending this approach to multiple long-lasting drug treatments, this study provides better insights into the factors affecting adherence and allows for a better identification of patients at high risk of non-adherence in drug taking.

PrgmNr 2327 - Higher adisposity, depression and wellbeing: causal inference using Mendelian randomization

View session detail

Author Block: **F. Casanova**[1], J. O'Loughlin[1], S. Martin[1], R. N. Beaumont[1], A. R. Wood[1], E. Watkins[2], R. Freathy[1], H. Saskia[3], T. Frayling[4], H. Yaghootkar[5], J. Tyrrell[4]; [1]Coll. of Med. and Hlth., Univ. of Exeter, Exeter, United Kingdom, [2]Sch. of Psychology, Univ. of Exeter, Exeter, United Kingdom, [3]Social, Genetic and Dev.al Psychiatry Ctr., Inst. of Psychiatry, Psychology and NeuroSci., King's Coll. London, London, United Kingdom, [4]Univ. of Exeter, Exeter, Devon, United Kingdom, [5]Univ. of Exeter Med. Sch., Exeter, Devon, United Kingdom

Disclosure Block: **F. Casanova:** None.

Higher adiposity is associated with higher odds of psychiatric diseases, including depression. These associations may be explained by the metabolic consequences and/or by the psychosocial impact of higher adiposity. Genetic variants associated with higher adiposity and with a favourable (lower triglycerides, higher HDL and lower type 2 diabetes risk) and unfavourable (higher triglycerides, lower HDL and higher type 2 diabetes risk) metabolic profile have been identified and can be used to understand which component of higher adiposity causes the highest risk.

We performed two-sample Mendelian Randomisation (MR) in up to 145,668 participants of European descent from the UK Biobank to test for a causal effect of higher adiposity on depression and wellbeing outcomes from the Mental Health Questionnaire (MHQ). We used three sets of adiposity genetic instruments: a) a set of 72 BMI genetic variants, b) a set of 36 favourable adiposity variants and c) a set of 38 unfavourable adiposity variants. We additionally tested causal relationships in a subset of individuals not taking antidepressants and using non-linear MR models.

Two-sample MR provided evidence that a genetically determined one standard deviation (1-SD) higher BMI (4.6â□□kg/m2) was associated with higher odds of depression [OR:1.50, 95%CI: 1.15, 1.95] and lower wellbeing [beta:-0.15, 95%CI:-0.26, -0.04]. When using the favourable and unfavourable adiposity variants higher adiposity was consistently associated with higher odds of depression (favourable adiposity OR:2.85, 95% CI:1.38, 5.89; unfavourable adiposity OR: 2.59, 95% CI:1.60, 4.21) and lower wellbeing scores (favourable adiposity beta:-0.20, 95%CI:-0.40, -0.01; unfavourable adiposity beta:-0.19, 95%CI:-0.39, 0.00). Excluding individuals on antidepressants had no effect on our results. A non-linear relationship was found between BMI and wellbeing: for individuals in the lowest BMI decile (32.5 kg/m2) a unit higher BMI was associated with lower wellbeing [Ã□:-0.12, 95%CI:-0.19, -0.05].

Our study provides further evidence that higher BMI is causally associated with higher odds of depression and lowers wellbeing. Using genetics to separate out metabolic and psychosocial effects, our study suggests that in the absence of adverse metabolic effects higher adiposity remains causal to depression and lowers wellbeing, highlighting the importance of psychosocial factors in this relationship. Our results showed evidence of non-linear causal association between adiposity and wellbeing.

PrgmNr 2328 - HLA Fine Mapping and Genetic Risk Factors in Tuberculosis

View session detail

Author Block: **A. Tervi**[1], N. Junna[1], M. Broberg[1], S. E. Jones[1], F. FinnGen[2], S. Strausz[1], C. A. Heckman[3], H. M. Ollila[1,4]; [1]Inst. for Molecular Med. Finland, Univ. of Helsinki, Helsinki, Finland, [2]FinnGen, Helsinki, Finland, [3]Inst. for Molecular Med. Finland, , HiLIFE-Helsinki Inst. of Life Sci., iCAN Digital Precision Cancer Med. Flagship, Univ. of Helsinki, Helsinki, Finland, [4]Broad Inst. of Harvard and MIT, Cambridge, MA

Disclosure Block: **A. Tervi:** None.

Background: Tuberculosis is a significant public health concern resulting in the death of over 1 million individuals each year worldwide. Smoking, BMI, HIV, diabetes and COPD have been suggested as risk factors. While treatment options and vaccines exist, a substantial number of infections still remain untreated or are caused by treatment resistant strains. Therefore, it is important to identify mechanisms that contribute to risk and prognosis of tuberculosis as this may provide tools to first understand disease mechanisms and provide novel treatment options for those with severe infection.
Aim: Our goal is to identify genetic risk factors that contribute to tuberculosis and to understand biological mechanisms and causality behind tuberculosis progression.
Results: A total of 1,895 individuals in FinnGen had ICD-based tuberculosis diagnosis. GWAS analysis identified three genetic variants with statistically significant association with tuberculosis at the Human leukocyte antigen (HLA) region (pINPP5A. Fine mapping the HLA-association provided evidence for one protective haplotype tagged by *DQB1*05:01* (p=1.82e-06, OR = 0.81 [CI 95% 0.74-0.88]), and predisposing alleles tagged by *HLA DQB1*06:04* (p=0.00030, OR = 1.34 [CI 95% 1.15-1.56]). Furthermore, genetic correlation analysis shows association with earlier reported risk factors including smoking and low BMI (pConclusions: Our findings indicate that genetic variants at the HLA region and *INPP5A* associate with the risk of tuberculosis. These findings suggest that the risk for tuberculosis is mediated by individual genetic variants and differences in the immune system. In addition, lifestyle risk factors such as smoking contribute to the risk of developing tuberculosis.

PrgmNr 2329 - Secondary findings from exome sequencing of 15,000 South Asians

View session detail

Author Block: **W. McLaren**[1], N. Sarkar Roy[2], Z. Nagda[2], N. Shaikh[2], S. Tuna[1], N. England[1], A. Shah[3], Regeneron Genetics Center, M. Kapoor[4], Y. Turpaz[5], J. D. Picker[6], A. R. Shuldiner[4], S. S. Jamuar[5]; [1]Global Gene Corp Pvt Ltd, Cambridge, United Kingdom, [2]Global Gene Corp Pvt Ltd, Mumbai, India, [3]Global Gene Corp Pvt Ltd, Ahmedabad, India, [4]Regeneron Genetics Ctr., Tarrytown, NY, [5]Global Gene Corp Pvt Ltd, Singapore, Singapore, [6]Global Gene Corp Pvt Ltd, Boston, MA
Disclosure Block: **W. McLaren:** None.
Background: In Western cohorts, the prevalence of secondary findings (SFs), referring to disease-causing variants in genes unrelated to a patient's primary condition or in healthy population-based sequencing programs, is between 0.86% and 8.8%. However, data on prevalence and type of SFs in South Asian (SA) populations is lacking. Moreover, genomic data from SA populations is underrepresented in public databases.
Methods: As part of our ongoing Indian biobank project we have collected and performed exome sequencing (ES) on leukocyte derived DNA from 15,185 individuals of SA origin. The GGC biobank includes both healthy individuals and a wide spectrum of disease cohorts, from diabetes, cardiovascular, neurodegenerative, oncology and others, and serves to build a longitudinal cohort for each disease area. We extracted and annotated variants from ES data in 59 ACMG-recommended genes (ACMG SF2.0) and filtered them to identify disease-causing variants based on ACMG-AMP criteria: pathogenic (P) or likely pathogenic (LP). We analyzed the distribution of SFs, class of gene, related medical conditions, and potential clinical impact.
Results: We found a total of 15,388 variants (14,762 SNVs, 576 in/dels) in the 59 genes in our cohort. Of these, 54% (8,326) are observed in only one individual and 38% (5,836) are novel variants not previously observed in public datasets (1000 Genomes, gnomAD). Among the novel variants, 194 (3.3%) are predicted to be loss-of-function. Filtering for rare and coding variants, we identified 253 potential disease causing variants present in 577 (3.8%) individuals. After reviewing family history and the ClinVar database, 148 (58.5%) of 253 variants were classified to be pathogenic, while an additional 105 (41.5%) variants were classified as LP. The overall cumulative prevalence of SFs (P and LP in ClinVar) in our cohort was 3.8% and could be as high as 5.1% including novel loss-of-functions.
Conclusion: The cumulative prevalence of SFs in our data was marginally higher than previously reported in Caucasian and South East Asian studies. Our ability to recontact individuals and use personal and/or family history to assess significance of the genomic findings demonstrates the value of longitudinal data collection in our biobank. More importantly, across 15,185 unrelated SA individuals, we identified 38% novel variants across the 59 genes, supporting the case for sequencing underrepresented populations to support precision medicine efforts globally.

PrgmNr 2330 - Separating the effects of cardiometabolic risk factors from type 2 diabetes on coronary and peripheral artery disease

View session detail

Author Block: **V. Walker**[1], M. Vujkovic[2], A. Carter[1], N. Davies[1], M. S. Udler[3], G. Davey Smith[4], B. F. Voight[2], T. Gaunt[1], S. M. Damrauer[5]; [1]Univ. of Bristol, Bristol, United Kingdom, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Broad Inst., Cambridge, MA, [4]Bristol Univ., Bristol, United Kingdom, [5]Hosp. of the Univ. of Pennsylvania, Philadelphia, PA
Disclosure Block: **V. Walker:** None.
Background: Type 2 diabetes and atherosclerotic cardiovascular disease share several risk factors. Yet, it is not fully understood if the effect of these risk factors on liability to atherosclerotic cardiovascular disease is mediated by liability to type 2 diabetes.
Methods: We performed univariate Mendelian randomization to quantify the effects of 57 continuous risk factors, selected for their clinical relevance to type 2 diabetes, on three outcomes: liability to type 2 diabetes, liability to coronary artery disease, and liability to peripheral artery disease. Next, we performed univariate Mendelian randomization to determine the effects of liability to type 2 diabetes on the risk factors, excluding risk factors with bidirectional relationship with liability to type 2 diabetes as they are unsuitable for mediation analyses. Finally, we performed two-step Mendelian randomization for mediation to estimate the mediating pathways between the risk factors, liability to type 2 diabetes, and liability to the atherosclerotic cardiovascular disease outcomes.
Results: Among the 57 risk factors, five were causes of liability to type 2 diabetes, 22 were causes of liability to coronary artery disease, and 22 were causes of liability to peripheral artery disease. A further 14 risk factors were consequences of liability to type 2 diabetes. We removed 16 risk factors, which had a bidirectional relationship with liability to type 2 diabetes, prior to conducting the mediation analyses. The effects of total cholesterol, diastolic blood pressure, hip circumference were likely to have direct effects on liability to both atherosclerotic cardiovascular disease outcomes. Height was also likely to have a direct effect on liability to coronary artery disease. These effects were unlikely to be mediated via liability to type 2 diabetes.
Conclusions: Separating the effects of risk factors from liability to type 2 diabetes can aid understanding of their relationships with liability to atherosclerotic cardiovascular disease. Our analysis suggests that the effects of these risk factors on liability to atherosclerotic cardiovascular disease are unlikely to be mediated by liability to type 2 diabetes. Therefore, control of the modifiable risk factors remains important for reducing atherosclerotic cardiovascular disease risk regardless of patient liability to type 2 diabetes.

PrgmNr 2331 - Transcriptome and GWAS identified rs2238678 as a critical SNP for eosinophilia in primary biliary cholangitis

View session detail

Author Block: **K. Ueno**[1], Y. Hitomi[2], Y. AIBA[3], O. Gervais[4], Y. Kawai[5], S-S. Khor[6], M. Kawashima[7], N. Nishida[8], K. Kojima[9], M. Nagasaki[10], K. Tokunaga[5], M. NAKAMURA[11]; [1]Res. Inst. Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [2]Hoshi Univ., Tokyo, Japan, [3]Omura, Japan, [4]Nihon Univ., Mishima, Japan, [5]Natl. Ctr. for Global Hlth.and Med., Tokyo, Japan, [6]Natl. Ctr. for Global Hlth.and Med., Tokyo, Tokyo, Japan, [7]Japan Sci. and Technology Agency, Chiyoda-ku, Japan, [8]Natl. Ctr. for Global Hlth.and Med., Ichikawa, Chiba, Japan, [9]Tohoku Univ, Sendai, Japan, [10]Kyoto Univ., Kyoto-City, Japan, [11]Natl. Hosp. Organization (NHO) Nagasaki Med. Ctr., Omura, Nagasaki, Japan
Disclosure Block: **K. Ueno:** None.
Primary biliary cholangitis (PBC) is a chronic cholestatic liver disease characterized by the destruction of biliary epithelial cells, presumably by autoimmune mechanism. Our recent transcriptome analysis of the diseased liver has identified that mRNA expression associated with eosinophilia is specifically increased in PBC as compared to chronic hepatitis C (CHC). The aim of this study is to investigate the molecular mechanisms of eosinophilia in PBC using GWAS and transcriptome datasets. By mRNA-microarray analysis of liver-biopsy samples (PBC: n=43, CHC: n=15), we firstly identified 256 genes that are significantly increased (Fold change>2, P

PrgmNr 2332 - Development of a novel, instrument-free, single-cell RNA sequencing technology (PIPseq) and its application to drug pathway discovery in lung cancer

View session detail

Author Block: **K. Fontanez**[1], I. Clark[2], C. Dâ□□Amato[1], A. Osman[1], S. Pandey[1], Y. Xue[1], A. May-Zhang[1], R. Meltzer[1], S. Kiani[1], A. Abate[3]; [1]Fluent BioSci., Watertown, MA, [2]Berkeley, Berkeley, CA, [3]Univ. of California, San Francisco, San Francisco, CA
Disclosure Block: **K. Fontanez:** Salary/Employment; Fluent BioSciences.
Single-cell RNA sequencing (scRNA-Seq) has enabled unprecedented insight into the biology of individual cells across a broad range of discovery and disease-relevant applications. Traditional scRNA-Seq workflows have included single cell sorting into wells, co-capture of cells with barcoded beads using microfluidics, or in-cell combinatorial indexing. Fluent BioSciences has developed Pre-templated Instant Partitions (PIPs) to simultaneously segregate complex cell mixtures into partitions with barcoded template particles that can be easily processed for scRNA-seq (PIPseq) without the need for complex instrumentation or microfluidic consumables. Here, we used PIPseq to bioinformatically discriminate the transcriptomes of Gefitinib resistant and sensitive cell lines after drug treatment. To evaluate tyrosine kinase inhibitor effects on adenocarcinoma cellular transcriptomics, PC9 cells, H1975 cells, or a mixed population (9:1) of PC9 and H1975, cells were treated with Gefitinib and processed with PIPseq. UMAP analyses of the transcriptome were performed and overlapped for each experiment. We demonstrate drug-treatment dependent gene expression changes in lung adenocarcinoma cells treated with the tyrosine kinase inhibitor Gefitinib. The resulting cell expression profiles clearly segregate by treatment condition in UMAP projections indicating PIPseqâ□□s ability to faithfully detect responses to drug exposure. These studies establish the efficacy of PIPseq for single-cell transcriptomic analysis across multiple drug treatment conditions. The simple PIPseq workflow is optimized for comparing multiple sample treatment conditions in a single controlled experiment. With minimal upfront cost of implementation, PIPseq can be easily implemented in any molecular research lab.

PrgmNr 2333 - Diversity of germline mutations in a cohort of unselected Brazilian pancreatic carcinoma patients

View session detail

Author Block: **L. Rodrigues**[1], S. Maistro[2], L. S. Leite[2], U. Ribeiro Jr[1], R. S. C. Guindalini[3], M. A. K. Folgueira[1]; [1]Faculdade de Med. da Univ.e de SÃ£o Paulo, SÃ£o Paulo, Brazil, [2]Inst. do CÃ¢ncer do Estado de SÃ£o Paulo, SÃ£o Paulo, Brazil, [3]Oncologia D'Or, SÃ£o Paulo, Brazil
Disclosure Block: **L. Rodrigues:** None.
Background: Only a limited number of pancreatic carcinoma patients have been tested for germline mutations in a large panel of genes, particularly outside the North Hemisphere. Therefore, we evaluated a cohort of unselected Brazilian pancreatic carcinoma patients, that represents a multiethnic population, to detect the spectrum and frequency of germline mutations in cancer predisposing genes. **Methods**: Patients from Instituto do CÃ¢ncer do Estado de SÃ£o Paulo, SÃ£o Paulo, Brazil, with histopathological diagnosis of non-endocrine pancreatic carcinoma were included, regardless of the family history of cancer. Genomic DNA was obtained from peripheral blood for Next Generation Sequencing (NextSeq platform) using the TruSight Hereditary Cancer panel, comprising 113 cancer predisposing genes. Variant analysis was performed with the VarStation, a Brazilian tool, that offers post-sequencing computational support. Reported variants are exonic, had coverage â¥150x, allelic frequency â¥0.3 and respect Clinvar deposits. **Results**: A total of 82 patients were evaluated, with mean age 61 years (34-87), among whom, 18% with young age (â¤50 years) and 48 women (57%). Forty-eight patients (57%) reported cases of any cancer in first-degree relatives. Regarding risk factors, 46 patients (55%) reported smoking, 18 alcohol abuse, 52 (62%) self-declared white and 28 were previously obese. Forty-five patients (54%) were diagnosed with T4 or metastatic disease. Sixteen out of 82 (19%) patients presented pathogenic/likely pathogenic variants in the following genes: *ATM, BRCA1, BRIP1, MTIF, MUTYH* (two patients in each gene) and *CDKN2A, CHEK2, FANCE, FANCL, MSH2, RAD51C, SDHA* and *SPINK1* (one patient in each gene), among these, two patients had two variants. **Conclusions**: In this unselected and multiethnic group of pancreatic carcinoma patients, 60% reported family history of cancer and 20% were carriers of pathogenic variants in cancer predisposing genes, mainly genes related with the homologous recombination DNA repair. Future analyzes in a greater number of patients will allow a better understanding of the mutational profile of the pancreatic carcinoma, especially in different populations.

PrgmNr 2334 - Effect of return of unknown significance on surgical variant of decision making in women with invasive breast cancer

View session detail

Author Block: **R. E. Ellsworth**[1], A. Vargason[2], C. E. Turner[3], C. D. Shriver[4]; [1]Murtha Cancer Ctr., Windber, PA, [2]Walter Reed Natl. Military Med. Ctr., Bethesda, MD, [3]Uniformed Services Univ. of the Hlth.Sci., Bethesda, MD, [4]Uniformed Services Univ. of the Hlth.Sci. and Walter Reed Natl. Military Med. Ctr., Bethesda, MD

Disclosure Block: **R.E. Ellsworth:** None.

Background While clinical management should not be influenced by a return of a variant of uncertain significance (VUS), uncertainty may influence surgical decision making of women with invasive breast cancer. We thus evaluated whether surgical choices differed between women with pathogenic, VUS and benign genetic test results. Methods Germline test results and all surgical procedures were extracted for women who had clinical testing within one year of diagnosis (n=620). Results were classified as pathogenic/likely pathogenic (15.8%), VUS (18.4%) or benign/likely benign (65.8%). Data were analyzed using chi-square tests with p

PrgmNr 2335 - Exome sequencing of 500 Brazilian patients with rare diseases: what we have learned

View session detail

Author Block: **C. D. Quaio**[1], C. Chung[2], S. Perazzio[2], A. Dutra[2], C. Moreira[2], C. Kim[3]; [1]Genetics Unit, Faculdade de Med. da universidade de Sao Paulo, Sao Paulo, Brazil, [2]Fleury Med. e Saude, Sao Paulo, Brazil, [3]Genetics Unit, Faculdade de Med. da Univ.e de Sao Paulo, Sao Paulo, Brazil
Disclosure Block: **C.D. Quaio:** Salary/Employment; Fleury Medicina e Saude.
Literature regarding genomic studies for rare diseases is limited in Latin America. We describe the variable spectrum of primary findings, secondary findings, carrier status for recessive diseases and clinical impact of exome sequencing (ES) in a cohort of 500 Brazilian patients with rare diseases. In total, 164 primary findings were reported in 158 patients (six patients presented dual diagnoses), representing an overall diagnostic yield of 31.6%. Most of the findings (61.6%) corresponded to autosomal dominant conditions, followed by autosomal recessive (25.6%) and X-linked (12.8%) conditions. The 158 patients found with primary findings had on their ES report 195 variants in total, among which 43.6% are novel in the literature. The rate of molecular diagnosis was considerably higher for prenatal samples (67%; 4/6), younger children (44%; 24/55), individuals with consanguinity (50%; 3/6), gastrointestinal/liver disease (44%; 16/36) and syndromic/malformative conditions (41%; 72/175). We observed a potential impact on the clinical care of the patients molecularly diagnosed in our study through the use of targeted therapy, early tumor screening, medication adjustment and monitoring for disease-specific complications in 15.6% of the patients in our cohort. Secondary findings were reported in 37 patients (7.4%). Based on cost-effectiveness studies in the literature, we speculate that the reports of secondary findings may influence an increase of 123.2 years in the life expectancy for our cohort, or 0.246 years/cohort patient. Additionally, carrier status for autosomal recessive diseases was studied in 320 patients: a total of 425 occurrences of 351 rare variants were reported in 278 different genes in 230 patients (71.9%); almost half (48.8%) of these 320 patients were carriers of at least one heterozygous pathogenic/likely pathogenic variant for rare metabolic disorders, while 25.9% of epilepsy, 18.1% of intellectual disabilities, 15.6% of skeletal disorders, 10.9% immune disorders, and 9.1% of hearing loss. In one of the largest cohorts of rare diseases in Latin America, we observed that ES is a powerful method to identify the molecular bases of monogenic disorders and redirect clinical care.

PrgmNr 2336 - Human genetics isgreat:Incidental karyotype finding of aBRCA1translocation explaining family cancer history

View session detail

Author Block: **M. RÃ©da**[1], C. Jacquot-Sawka[1], J. Sokolowska[2], C. Bonnet[2], N. Marle[3], M. Bronner[2], B. Leotard[2], S. Cacciatore[2], A. Baurand[1], P. Callier[4], A. Saunier[2], P. Jonveaux[2], A. Vitobello[5,6], M. Chevarin[5], Y. Duffourd[5], L. Faivre[1], S. Nambot[1]; [1]Genetic Ctr., CHRU Dijon, Dijon, France, [2]Genetic Lab. and INSERM U-954, CHRU Nancy, Nancy, France, [3]Cytogenetic Lab., CHRU Dijon, Dijon, France, [4]Genetic Lab., CHRU Dijon, Dijon, France, [5]INSERM UMR 1231 GAD, Univ. of Burgundy, Dijon, France, [6]UF 'Genomic diagnosis innovation in rare diseases', CHRU Dijon, Dijon, France
Disclosure Block: **M. RÃ©da:** None.
Background Molecular study of BRCA1/2 genes in search for cancer predisposition is currently recommended based on patientsâ age at the diagnosis and/or family history of breast or ovarian cancer. Approximately 10-15% of cases are associated with BRCA germline variants. A few cases are explained by recent predisposition genes, but a majority of cases remain unsolved. This could be related to the impact of unknown genes, but also to a failure in detecting variations by current sequencing techniques. This report illustrates the latest hypothesis by highlighting a genetic predisposition through an unusual technique in a HBOC family. **Case report** We report here the case of a female patient who presented a breast cancer at 46 years-old. Her personal history included in-utero fetal death at 8.5 months, an extra-uterine pregnancy and two spontaneous miscarriages. Her mother also presented with breast cancer at 41 years-old and her maternal grandmother presented with breast cancer at 48 years-old. Amongst her 5 siblings, one of her sisters presented with ovarian cancer at 37 years-old and breast cancer at 43 years-old. The characteristic association of several cases of breast and ovarian cancer in first-degree relatives led to BRCA1/2 genes analysis at the beginning of the 2000s in the two sisters. Both tests were negative. An antenatal karyotype was realized during her sisterâs pregnancy, because of elevated serum markers of Down syndrome. It revealed a translocation t(1;17)(q41;q21.31), also present in our patient. Array-CGH did not show imbalance at break points. As the interested region of chromosome 17 could concern the BRCA1 gene, a FISH analysis was carried out to identify breakpoints: breakpoint was located at BAC RP11-812O05 (intronic) and included the BRCA1 gene. A RT-qPCR was then performed on the circulating RNA of one of the patients and showed a decrease of BRCA1 gene expression. This functional analysis demonstrated the impact of the translocation on BRCA1 expression, allowing predictive tests in the family through karyotype. **Discussion** Alterations in the BRCA1 gene are mainly single nucleotide variants (accounting for 10% to 50% of the germline mutations) and large-scale rearrangements. Balanced translocations are rarely searched for and are not always detectable by gene panel analysis using short reads technic. **Conclusion** We report here an incidental karyotype finding of BRCA1 predisposition to breast/ovarian cancers that reveals the lacks of detection of classical techniques for certain variants and shows the interest of whole genome sequencing in unsolved HBOC families. Further analysis such as genome study and bionano are ongoing in this case.

PrgmNr 2337 - Searching for epigenetic markers with prognostic value in Acute Myeloblastic Leukemia patients of Uruguay

View session detail

Author Block: **M. Cappetta**[1], S. Pereyra[1], N. Delloca[1], F. Salvarrey[1], R. Neumann[2], C. May[2], V. Elizondo[3,4], G. Manrique[3,4], V. PÃ©rez[4], B. Bertoni[1], M. Zubillaga[3,4]; [1]Departamento de GenÃ©tica, Facultad de Med., Univ. de la RepÃºblica, Montevideo, Uruguay, [2]Dept. of Genetics, Coll. of Med., Univ. of Leicester, Leicester, United Kingdom, [3]Laboratorio de BiologÃa Molecular, MÃ©dica Uruguaya CorporaciÃ³n de Asistencia MÃ©dica (MUCAM), Montevideo, Uruguay, [4]Laboratorio de BiologÃa Molecular, AsociaciÃ³n EspaÃ±ola Primera en Salud, Montevideo, Uruguay
Disclosure Block: **M. Cappetta:** None.
Acute myeloblastic leukemias (AML) are the most frequent acute leukemias in adults. Genomic and candidate gene studies identified novel recurrent somatic genetic and epigenetic variants in AML patients with biological, clinical and therapeutic significance. In Uruguay, the characterization of AML is dependent on integration of cytomorphology, immunophenotype, cytogenetics and molecular biology (*FLT3*, *NPM1*, *CEBPA*, *cKIT* mutations). This allows stratification in prognostic risk groups and rationalization of therapeutic resources. However, some patients with normal karyotype remain unstratified due to failed detection of available markers.
In order to expand the AML molecular markers analysis in Uruguayan patients, we evaluated methylation levels in promoters of 13 candidate genes in bone marrow or peripheral blood DNA samples from 37 adult AML patients at the onset of the disease and 3 healthy controls. Gene promoters were amplified from sodium bisulfite treated DNA and sequenced using MinION platform (Nanopore). A bioinformatic analysis pipeline was designed by our research group to determine methylation status of each CpG in the promoter regions studied.
In this work we detected hypermethylated regions within the gene promoters of *DNMT3A*, *IDH1*, *FLT3* and *JAK3* in AML patients compared to healthy controls. On the other hand, in the *TP53*, *NRAS*, *WT1*, *ASXL1* and *SH2B3* gene promoters specific differentially methylated CpG sites are described in AML patients (Wilcoxon test, p FLT3, *DNMT3A*, *IDH1*, *TP53*, *JAK3* and *NRAS* show differential methylation in AML patients with an extremely acute manifestation of the disease (with a rapidly unfavorable evolution) compared to the rest of the patients analyzed (Wilcoxon test, p We report for the first time an epigenetic analysis in Uruguayan AML patients. These results will be validated in a larger sample and analyzed in conjunction with the clinical and paraclinical data, in order to study the relevance of these markers. However, this study highlights the relevance of detecting epigenetic variability underlying prognostic risk subgroups to deliver patient tailored clinical decision support.

PrgmNr 2338 - Translocation t(8;14)(q11;q32) in Leukemia, a literature review

View session detail

Author Block: **Y. Llimpe Mitma de Barron**; UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS, Lima, Peru
Disclosure Block: **Y. Llimpe Mitma de Barron:** None.
The translocation t(8;14)(q11;q32) is an abnormality that affects *CEBPD* (8q11) and *IGH* (14q32) genes; and is defined as a distinct cytogenetic entity of precursor B ALL. This study compiles all the cases with t(8;14)(q11;q32) found in the Mitelman database (https://mitelmandatabase.isb-cgc.org/). It was found a total of 76 cases: 44 male and 32 female patients with ages ranged from 3 to 64 years-old (average age: 16.9). The leukocyte count (WBC) ranged from 0.7 to 172.1x10^9/L (average WBC: 25.49x10^9/L). The most frequent diagnosis was ALL; however, a case of AML and another of CML were also found. 21 patients presented the t(8;14)(q11;q32) as the only detectable abnormality; 12 cases presented the der(14)t(8;14)(q11;q32); 25 cases with constitutional trisomy 21 (Down Syndrome) and 6 with acquired trisomy 21. Most abnormalities associated with the t(8;14)(q11;q32) were the constitutional trisomy 21 (33%), followed by gain of X chromosome (18%), t(9;22)(q34;q11)(12%) and acquired trisomy 21 (8%). Other abnormalities found in this study include i(17)(q10)(n=4), +5(n=4), +4(n=3). In conclusion as it is described in the literature review, the t(8;14)(q11;q32) is a non-common abnormality and is associated with B-cell phenotype and often present high-risk features.

PrgmNr 2339 - *BRCA1* and *BRCA2* variants of unclear significance in hereditary breast and ovarian cancer patients. Looking for novel pathogenic variants

View session detail

Author Block: **A. Fiorino**[1], D. Fanale[1], L. Incorvaia[2], C. Filorizzo[1], A. Dimino[1], C. Brando[1], Z. Anwar[1], N. Barraco[1], M. Bono[1], K. Calcara[1], V. CalÃ²[1], D. Cancelliere[1], L. Corsini[1], A. Cucinella[1], S. Cutaia[1], V. Gennusa[1], M. Greco[1], V. Gristina[1], G. Madonia[1], L. Magrin[1], E. Pedone[1], A. Perez[1], A. Pivetti[1], M. Ricciardi[1], S. Sammataro[1], R. Scalia[1], R. Sciacchitano[1], G. Badalamenti[1], V. Bazan[2], A. Russo[1]; [1]Dept. of Surgical, Oncological and Oral Sci., Section of Med. Oncology, Univ. of Palermo, Palermo, Italy, [2]Dept. of Biomedicine, NeuroSci. and Advanced Diagnostics (Bi.N.D.), Section of Med. Oncology, Univ. of Palermo, Palermo, Italy

Disclosure Block: **A. Fiorino:** None.

About 10-20% of hereditary breast and/or ovarian (HBOC) cancer patients undergoing germline *BRCA1/2* genetic testing harbour Variants of Uncertain Significance (VUSs). Poor is the knowledge about the prevalence of germline *BRCA1/2* VUSs in HBOC patients of Southern Italy. Our study is aimed at describing the spectrum of these variants detected in HBOC patients in order to improve the patientâ□□s stratification with the identification of potentially high-risk *BRCA1/2* variants helpful for patient clinical management. 874 breast (BC) or ovarian (OC) cancer patients, enrolled from October 2016 to April 2021 at the â□□Sicilian Regional Center for the Prevention, Diagnosis and Treatment of Rare and Heredo-Familial Tumorsâ□□ of University Hospital Policlinico â□□P. Giacconeâ□□ of Palermo, were genetically tested for germline *BRCA1/2* variants through Next-Generation Sequencing analysis. The screening results showed that 639 (73.1%) out of 874 patients were *BRCA*-wild-type, whereas 67 (7.7%) were carriers of germline *BRCA1/2* VUSs and 168 (19.2%) harboured germline *BRCA1/2* Pathogenic/Likely Pathogenic Variants. Overall, the mutational analysis revealed the presence of 59 different VUSs detected in 67 patients, 46 of which affected by BC and 21 by OC. Twenty-one (35.6%) out of 59 variants were located on *BRCA1* gene, whereas 38 (64.4%) on *BRCA2*. We have identified six alterations in *BRCA1* and two in *BRCA2* with unclear interpretation of clinical significance. Familial anamnesis of a patient harbouring *BRCA1*-c.3367G>T suggests for this variant a potential of pathogenicity as well as the *BRCA1*-c.4963T>G variant identified in three unrelated OC patients. Understanding clinical significance of germline *BRCA1/2* VUSs could improve the identification of potentially high-risk variants useful for clinical management of BC/OC patients and family members. Reclassifying these variants could make them useful for predictive, prognostic and preventive purposes in clinical practice. Advances in molecular biology, such as the use of multi-gene panels, exome sequencing and/or RNA-seq, are increasing the amount of data in the field of research about VUSs. Further linkage analyses will be able in the future to provide additional information useful for understanding inherited variants of unclear clinical significance associated with HBOC.

PrgmNr 2340 - Breast and prostate cancer risks for male *BRCA1* and *BRCA2* pathogenic variant carriers using recently-developed polygenic risk scores

View session detail

Author Block: **D. Barnes**[1], V. Silvestri[2], G. Leslie[1], R. K. Schmutzler[3,4,5], S. L. Neuhausen[6], D. F. Easton[1,7], G. Chenevix-Trench[8], A. C. Antoniou[1], L. Ottini[2], Consortium of Investigators of Modifiers of BRCA1/2; [1]Ctr. for Cancer Genetic Epidemiology, Dept. of Publ. Hlth.& Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, [2]Dept. of Molecular Med., Univ. La Sapienza, Rome, Italy, [3]Ctr. for Familial Breast and Ovarian Cancer, Faculty of Med. and Univ. Hosp. Cologne, Univ. of Cologne, Cologne, Germany, [4]Ctr. for Integrated Oncology (CIO), Faculty of Med. and Univ. Hosp. Cologne, Univ. of Cologne, Cologne, Germany, [5]Ctr. for Molecular Med. Cologne (CMMC), Faculty of Med. and Univ. Hosp. Cologne, Cologne, Germany, [6]Dept. of Population Sci., Beckman Res. Inst. of City of Hope, Duarte, CA, [7]Ctr. for Cancer Genetic Epidemiology, Dept. of Oncology, Univ. of Cambridge, Cambridge, United Kingdom, [8]Queensland Inst. of Med. Res., Brisbane, Australia

Disclosure Block: **D. Barnes:** None.

Recent population-based female breast cancer and prostate cancer polygenic risk scores (PRS) have been developed. We assessed the associations and implications on cancer risk prediction of these PRS with breast cancer and prostate cancer risks for male *BRCA1* and *BRCA2* pathogenic variant carriers. We analyzed data from 483 *BRCA1* and 1,318 *BRCA2* male carriers of European ancestry available through the Consortium of Investigators of Modifiers of *BRCA1* and *BRCA2* (CIMBA). A 147-single nucleotide polymorphism (SNP) prostate cancer PRS (PRS_{PC}) and a 313-SNP breast cancer PRS were evaluated. There were three versions of the breast cancer PRS, optimized to predict overall (PRS_{BC}), estrogen-receptor (ER) negative (PRS_{ER-}) or ER-positive (PRS_{ER+}) breast cancer risk. The PRS_{ER+} yielded the strongest association with breast cancer risk. The odds ratio (OR) per PRS_{ER+} standard deviation estimates were OR=1.40 (95%CI: 1.07-1.83) for *BRCA1* and OR=1.33 (95%CI: 1.16-1.52) for *BRCA2* carriers. The PRS_{PC} was associated with prostate cancer risk for both *BRCA1* (OR=1.73, 95%CI: 1.28-2.33) and *BRCA2* (OR=1.60, 95%CI: 1.34-1.91) carriers. The estimated breast cancer ORs were larger after adjusting for female relative breast cancer family history, whereas the breast and prostate cancer ORs, adjusted for family history of male breast and prostate cancer, respectively, were similar to unadjusted estimates. By age 85 years, for *BRCA2* carriers, the breast cancer risk varied from 8% to 18% and prostate cancer risk from 34% to 88% between the 5^{th} and 95^{th} percentiles of the PRS distributions. Population-based prostate and female breast cancer PRS are associated with a wide range of absolute breast and prostate cancer risks for male *BRCA1* and *BRCA2* carriers. These findings warrant further investigation aimed at providing personalized cancer risks for male carriers and to inform clinical management.

PrgmNr 2341 - Breast cancer risks associated with missense variants in breast cancer susceptibility genes

View session detail

Author Block: **L. Dorling**[1], S. Carvalho[1], J. Allen[1], M. T. Parsons[2], C. Fortuno[2], A. GonzÃ¡lez-Neira[3], P. Pharoah[4], M. K. Schmidt[5,6], J. Benitez[7], B. Vroling[8], A. M. Dunning[9], S. H. Teo[10], A. Kvist[11], M. de la Hoya[12], P. Devilee[13], A. B. Spurdle[14], M. P. Vreeswijk[13], D. F. Easton[1], The Breast Cancer Association Consortium; [1]Ctr. for Cancer Genetic Epidemiology, Univ. of Cambridge, Cambridge, United Kingdom, [2]Dept. of Genetics and Computational Biology, QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [3]Human Cancer Genetics Programme, Spanish Natl. Cancer Res. Ctr. (CNIO), Madrid, Spain, [4]Univ Cambridge, Cambridge, United Kingdom, [5]Div. of Molecular Pathology, The Netherlands Cancer Inst. - Antoni van Leeuwenhoek Hosp., Amsterdam, Netherlands, [6]Div. of Psychosocial Res. and Epidemiology, The Netherlands Cancer Inst. - Antoni van Leeuwenhoek Hosp., Amsterdam, Netherlands, [7]Human Genetics CNIO - Carlos III, Madrid, Madrid, Spain, [8]Bio-Prodict, Nijmegen, Netherlands, [9]Univ. of Cambridge, Cambridge, United Kingdom, [10]Cancer Res. Initiatives Fndn., Subang Jaya, Selangor, Malaysia, [11]Lund Univ., Lund, Sweden, [12]Molecular Oncology Lab., CIBERONC, Hosp. Clinico San Carlos, IdISSC, Madrid, Spain, [13]Leiden Univ Med Ctr, Leiden, Netherlands, [14]QIMR Berghofer Med. Res. Inst., Brisbane, Australia

Disclosure Block: **L. Dorling:** None.

Protein truncating variants (PTVs) in *ATM*, *BRCA1*, *BRCA2*, *CHEK2* and *PALB2* are associated with an increased risk of breast cancer, but the risks associated with missense variants in these genes are uncertain. With a total of 59,639 breast cancer cases and 53,165 controls from 44 studies including European and Asian women, we took a training (80%) and validation (20%) approach to analyse data on rare (frequency ATM (1,146 training variants), *BRCA1* (664), *BRCA2* (1,425), *CHEK2* (325) and *PALB2* (472). We evaluated risks according to five *in-silico* risk prediction algorithms (Align-GVGD, BayesDel, CADD, Helix and REVEL), functional protein domain and frequency. We evaluated breast cancer risks using empirical logistic regression models as well as mixture models in which a subset of variants were assumed to be risk associated. The most predictive *in-silico* algorithms were Helix (*BRCA1*, *BRCA2* and *CHEK2*) and CADD (*ATM*). Increased risks appeared restricted to functional protein domains for *ATM* (FAT and PIK domains) and *BRCA1* (RING and BRCT domains). For *ATM*, *BRCA1* and *BRCA2* the data were compatible with the majority of the risk coming from small subsets (approximately 7%, 2% and 0.6%, respectively) of rare missense variants conferring similar risk to those of PTVs in the same genes. In contrast, for *CHEK2*, the data were more consistent with a large fraction (approximately 60%) of rare missense variants conferring an increased risk but lower (OR 1.75, 95% CI (1.47-2.08)) than *CHEK2* PTVs. There was little evidence for an association with risk for missense variants in *PALB2*. The best fitting training models were well calibrated in the validation set. These results inform both the selection of candidate variants for functional assays and risk prediction models, and thus could contribute to the clinical reporting of results to women undergoing gene panel testing for breast cancer susceptibility.

PrgmNr 2342 - Copy number variation (CNV) analysis identifies variants in 1p36 in African American and Caucasian hereditary prostate cancer cases

View session detail

Author Block: **A. Williams**[1], K. W. Termine[1], J. Waldron[1], O. Sartor[2], J. E. Bailey-Wilson[3], D. M. Mandal[1]; [1]LSU Hlth Sci Ctr, Genetics, New Orleans, LA, [2]Tulane Med. Sch., New Orleans, LA, [3]NIH/NHGRI, Baltimore, MD

Disclosure Block: **A. Williams:** None.

Prostate cancer (PCa) is a common malignancy which affects 1 in 8 men. There is a significant racial disparity and African American (AA) males are more at risk for developing such cancers, at a rate of almost double compared to the males of European ancestry (EA). So far, not much is known about the role of germline copy number variations (CNVs) in this health disparity. Our previous work has shown several genetic regions with CNVs in both AA and EA hereditary prostate cancer (HPC) cases. The goal of this project was to detect germline CNVs in the targeted resequencing data spanning 9 Mb region in 1p36, that was previously identified by microarray and whole exome sequencing analyses. For this study, a total of 50 individuals were used, 25 AA and 25 EA men from HPC families. We have used three CNV calling algorithms: XHMM, CANOES, and GATK4. First, we focused on four PCa associated genes: *NBPF1*, *NBL1*, *SRSF10*, and *RHD*, that were previously identified in our study. In the current CNV analysis, XHMM identified deletions in *NBPF1* in several samples in both AA and EA cases. GATK4 was unable to call any CNVs in this gene. *NBL1* had no identified deletions in any tool, despite previous microarray data to the contrary. XHMM identified full deletions of SRSF10 in most of the cases, while GATK4 identified partial deletion in several cases from both ancestries, but more frequently in AA cases. Finally, a deletion was detected in *RHD* in only a few cases, and the deletion was confirmed by both XHMM and GATK4 algorithms. Deletion in *RHD* was more common in the EA population than the AA population. CANOES was unable to identify any variants within our regions of interest. We then expanded our search to other identified variants to identify regions commonly detected by the CNV calling algorithms. A region of deletion was detected in the *CELA3A* gene (reported to be downregulated in pancreatic and prostate cancer) and a region between *AKR7L* and *AKR7A3* both of which are found to be frequently mutated in cancer cells. XHMM called the deletion of both of these regions at a much higher rate than GATK4: nearly in all our cases from the two ancestries, compared to only a few in GATK4 and CANOES. The few PCa cases who had deletions across all tools were always in the AA population. Altogether, this new analytical strategy establishes the usefulness of applying multiple CNV callers in identifying regions of potential interest, as well as verifying the results from previous studies of PCa. Results from this study may hold valuable information in finding potential biomarkers to address PCa health disparity in the future. Further validation of the identified variants is ongoing.

PrgmNr 2343 - Genetically Predicted Serum Insulin-like Growth Factor-1 (IGF-1) Levels and Risk of Multiple Primary Cancers in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial

View session detail

Author Block: **S. He**, C. Breeze, W. Wong, M. Machiela, W-Y. Huang, N. Freedman, S. I. Berndt; Natl. Cancer Inst., Rockville, MD
Disclosure Block: **S. He:** None.
Objectives: Epidemiologic studies have reported positive associations between elevated levels of serum insulin-like growth factor-1 (IGF-1) and the risk of several common cancers, such as prostate and colorectal cancer, leading to speculation that IGF-1 may contribute to the risk of multiple primary cancers. Multiple primary cancers are estimated to occur in 2-17% of the population, but etiologic factors are not well understood. This study aims to investigate the causal relationship between the IGF-1 pathway and the risk of multiple primaries using a Mendelian randomization approach.
Methods: We conducted a trans-ancestry case-control study including 4,402 cases with multiple primaries and 76,419 controls from the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. Using a previously derived polygenic risk scores (PRS) for serum IGF-1 levels based on 23,443 SNPs , we calculated scores for each subject using PRSice2. This PRS is estimated to explain approximately 22% of the variance in Europeans, but less in other populations (Sinnott-Armstrong N, et al. 2021). We used logistic regression to estimate the odds ratios (ORs) for the association with multiple primaries, adjusting for age, sex, genotype platform and ancestry-specific principle components. The PRS was analyzed both as a continuous measure and as quintiles, based on the distribution among controls. Analyses were conducted separately for European (EUR), African American (AA), and East Asian (EA) populations. We then meta-analyzed the ancestry-specific results using a fixed effects model.
Results: No significant associations were observed between genetically predicted IGF-1 levels and the risk of multiple primaries in either ancestry-specific analyses or the trans-ancestry meta-analysis (trans-ancestry OR (95%CI): 1.18 (0.82, 1.70) for the IGF-1 PRS as a continuous measure; 1.03 (0.93, 1.13)) for the highest PRS quintile compared to the lowest one). **Conclusions:** Our trans-ancestry study does not support an association between genetically predicted serum IGF-1 levels and multiple primaries. Further polygenic risk score analyses are planned to explore the relationship between other biomarkers and the risk of multiple primaries.

PrgmNr 2344 - Germline *PALB2* mutations found in NGS multigenetic panels from breast-ovarian cancer patients in Argentina

View session detail

Author Block: **A. Gonzalez**[1], G. Tabares[2], A. Mampel[3], C. Montes[4], C. Martin[5], L. Vargas Roig[6], M. Lopez[7], N. Rossi[8], P. Quaglio[9], A. Yanzi[10], S. Acevedo[11], L. Lugo[12], P. Lopez[13], M. Del Castillo[4], M. Vazquez[14], L. NuÃ±ez[15]; [1]Heritas, Rosario, Argentina, [2]CEMA Centro de Mastologia Rosario, Rosario, Argentina, [3]Univ. Natl. de Cuyo. COIR, Mendoza, Argentina, [4]Inst. modelo de ginecologÃa y obstetricia, CÃ³rdoba, Argentina, [5]Hosp. Privado Univ.rio de CÃ³rdoba, CÃ³rdoba, Argentina, [6]IMBECU-CONICET-UNCuyo, Mendoza, Argentina, [7]CEMAFE, Santa Fe, Argentina, [8]Fundacion para el Progreso de la Med., CÃ³rdoba, Argentina, [9]Laboratorio Cigen, Rosario, Argentina, [10]Inst. San Marcos, San Juan, Argentina, [11]Hosp. BritÃ¡nico de Buenos Aires, Buenos Aires, Argentina, [12]Clinica San Geronimo, Santa Fe, Argentina, [13]Inst. de Ginecologia de Rosario, Rosario, Argentina, [14]Heritas - CONICET, Rosario, Argentina, [15]Hosp. Aleman de Buenos Aires, Buenos Aires, Argentina
Disclosure Block: **A. Gonzalez:** None.

During the last decade, *PALB2* has been included as one of the most relevant high-risk breast cancer predisposing genes after *BRCA1/2*. *PALB2* variants have been scarcely described in Argentinian reports, mainly in the context of multi genetic panel (MP) testing along with other hereditary cancer genes (Cerretini R, et al.2019). In this study, we describe molecular and clinical characteristics of *PALB2* mutations found in 1276 MP from breast-ovarian cancer families in different institutions from Argentina. PALB2 pathogenic (PV) and probably pathogenic (PPV) variants were retrospectively identified and MP reports collected from one local reference lab, Heritas and SITHER public database, from 2017 to present. Geneticists in charge of the patients were contacted to provide anonymous clinical data. All patients were adequately counseled and met NCCN criteria for hereditary breast/ovarian cancer (HBOC) testing. MP were performed by different laboratories with Illumina NGS technology and the number of genes per panel ranged from 5 to 30. All MP included actionable genes associated with HBOC and all the variants were reviewed using the ACMG guidelines. VUS and familial mutations were excluded from this report. A total of 32 *PALB2* PV and 4 PPV were identified. Three of the PPV are novel. The frequency of *PALB2* mutations in the MP cohort is 2,8% (36/1276) similar to other reports. Regarding type of mutations 60% are nonsense, 33% frameshift and 7% large rearrangement. We found 3 double heterozygotes with a PV in *PALB2* and also in *TP53, ATM* and *BRCA2* genes. One recurrent PV (c.1653T>A,p.Tyr551*) was present in 47% (17/36) of carriers. Although previously reported , such a high recurrence mutation rate justifies further research. All patients were unrelated index cases, 32 (89%) with breast cancer (BC), 1 (3%) prostate cancer, 1 (3%) pancreatic cancer and 1 with BC and ovarian cancer concomitant diagnosis. One patient was unaffected at the time of testing (positive family history). BC cases were mainly ductal carcinoma (76%) , luminal (57%) and triple-negative (24%). The mean age at BC diagnosis was 43.5 years (23-60) and there were 5 (15%) cases of bilateral BC and 4 (12%) cases of second BC. Ethnic background is mainly of western European origin (53% of patients with Spanish ancestry and 36% Italian) and less than 10% native Amerindian. To our knowledge, this is the first report describing molecular and clinical characteristics of PALB2 carriers in Argentina. The frequency of PALB2 PV in families with HBOC suspicion is similar to other populations and Tyr551* is a recurrent mutation that seems to be responsible for almost 50% of the cases.

PrgmNr 2345 - Germline genetically predicted mammographic density is associated with breast tumor transcriptomic and proteomic features, CD8+ T cell infiltration, somatic mutational signatures, and tumor mutational burden in The Cancer Genome Atlas

View session detail

Author Block: **A. Francis**[1,2], T. Robinson[1,2], B. Lloyd-Lewis[3], Y. Liu[4], C. Peng[4,5], R. Tamimi[6], P. Kraft[4], P. Brennan[7], T. Gaunt[1,2], R. M. Martin[1,2,8], C. Relton[1,2], A. Gabriel[9], G. Gierach[10], J. McKay[9], S. Kar[1,2]; [1]MRC Integrative Epidemiology Unit, Univ. of Bristol, Bristol, United Kingdom, [2]Population Hlth.Sci., Bristol Med. Sch., Univ. of Bristol, Bristol, United Kingdom, [3]Sch. of Cellular and Molecular Med., Univ. of Bristol, Bristol, United Kingdom, [4]Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [5]Channing Div. of Network Med., Brigham and Womenâ□□s Hosp., Boston, MA, [6]Div. of Epidemiology, Dept. of Population Hlth.Sci., Weill Cornell Med., New York, NY, [7]Genetic Epidemiology Group, Intl. Agency for Res. on Cancer, Lyon, France, [8]Univ. Hosp. Bristol and Weston NHS Fndn. Trust, NIHR Bristol BioMed. Res. Ctr., Univ. of Bristol, Bristol, United Kingdom, [9]Genetic Cancer Susceptibility Group, Intl. Agency for Res. on Cancer, Lyon, France, [10]Integrative Tumor Epidemiology Branch, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD

Disclosure Block: **A. Francis:** None.

High mammographic density (MD) is one of the strongest risk factors for breast cancer, the most common cancer in women globally, but the molecular underpinnings of this association are poorly understood. Tumor cohorts with multi-omic phenotyping currently do not have linked MD data at scale, making it challenging to define the genomic landscape of breast tumors arising on a background of high MD. We took a new approach to address this challenge by integrating MD genome-wide association study (GWAS) data with The Cancer Genome Atlas (TCGA) breast tumor data. We built a polygenic score for MD using effect size estimates and allele information for the lead variants at 17 independent loci (P -8) from a GWAS of percent MD in 24,192 women of European ancestry. We performed sample and genotype quality control and imputation into the 1000 Genomes reference panel on germline genetic data from 721 female breast cancer cases of genetically inferred European ancestry from TCGA. We assigned each woman in this TCGA cohort, which lacks measured MD, her genetically predicted MD based on the MD polygenic score. We evaluated the association between germline genetically predicted MD and breast tumor genome-wide transcriptomic, proteomic, genomic, epigenomic, and immune traits in TCGA using linear (default) and quasi-Poisson (for overdispersed count data) regression models adjusted for age and stage at diagnosis, estrogen receptor status, and 10 genetic principal components, reporting associations at FDR DLL1, an estrogen-dependent Notch ligand previously implicated in breast oncogenesis ($P = 2 \times 10^{-6}$; top gene of 20,530 genes profiled by RNA-Seq). High genetically predicted MD was associated with increased breast tumor CD8+ T cell infiltration (P = 0.004) in our evaluation of 22 tumor immune infiltrates profiled by the CIBERSORT algorithm. Genetically predicted MD had positive and inverse associations with breast tumor MAPK ($P = 3 \times 10^{-4}$) and XRCC1 ($P = 4 \times 10^{-4}$) protein levels, respectively, of the 281 tumor proteins profiled by reverse-phase protein array, suggesting roles for mitogen-activated protein kinase signaling and DNA repair. High genetically predicted MD was associated with the mitotic clock-like, aging-related single base substitution mutational signatures 1 (P = 0.001) and 5 (P = 0.002) and with high tumor mutational load (P = 0.006 for association with non-silent mutations/Mb). Thus, we combined germline and somatic data to identify breast tumor molecular features associated with genetically predicted MD, with potential implications for breast cancer development and progression.

PrgmNr 2346 - Identification of copy number variants in hereditary lung cancer families

View session detail

Author Block: **J. Waldron**[1], K. W. Termine[1], A. M. Musolf[2], M. de Andrade[3], C. Gaba[4], R. Govindan[5], P. Yang[3], M. You[6], M. W. Anderson[7], A. G. Schwartz[8], S. M. Pinney[7], C. I. Amos[9], J. E. Bailey-Wilson[2], D. M. Mandal[1]; [1]Dept. of Genetics, LSU Hlth.Sci. Ctr., New Orleans, LA, [2]Natl. Human Genome Res. Inst., NIH, Baltimore, MD, [3]Mayo Clinic, Rochester, MN, [4]Univ. of Toledo, Toledo, OH, [5]Div. of Oncology, Washington Univ. Sch. of Med., St. Louis, MO, [6]Med. Coll. of Wisconsin, Milwaukee, WI, [7]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [8]Karmanos Cancer Inst. Wayne State Univ., Detroit, MI, [9]Baylor Coll. of Med., Houston, TX

Disclosure Block: **J. Waldron:** None.

Lung cancer is the most common cause of cancer mortality and the third most common cancer by incidence in the United States. While environmental factors (e.g. tobacco smoke) play an important role in its development, lung cancer risk also exhibits a high degree of heritability. Linkage analyses and genome-wide association studies have identified multiple loci associated with increased lung cancer risk; however, much of the heritability has yet to be explained by these loci. Structural mutations, including a gain or loss of DNA (copy number variants or CNVs), contribute to phenotypic diversity through dosage and/or cis-regulatory effects. CNVs are important sources of phenotypic variation and likely contribute a larger fraction of genomic variation among individuals than SNPs. Unbiased CNV mapping is only possible with the use of high-depth, short-read sequencing and remains much more complicated and prone to false positives than SNP detection. As a result, the contribution of CNVs to genetic variation is not as well understood as the impact of SNPs. As the tools necessary for CNV detection have improved, both somatic and germline CNVs have been shown to play an important role in disease, especially cancer. Although CNVs have been shown to be common and often benign, they account for a significant proportion of pathogenic variants. Little has been done to understand the role of germline CNVs in the biological pathways of hereditary lung cancer. The goal of the current project is to utilize the whole exome sequencing (WES) data in identifying the CNVs in the hereditary lung cancer (HLC) families (â¥3 LC/family) recruited by the Genetic Epidemiology of Lung Cancer Consortium (GELCC). This work uses germline WES data from 203 individuals (60 with a lung cancer diagnosis) from 25 HLC families. We limited our investigation to CNVs called by two independent tools using read depth to infer copy number changes and genomic breakpoints: GATK 4 (https://gatk.broadinstitute.org) and XHMM (https://atgu.mgh.harvard.edu/xhmm/). Those CNVs that segregate with disease, are consistent with Mendelian expectations, and appear to be uncommon in the general population, are subjected to bioinformatic annotation to detect the most probable causal variants in each family. The enrichment of rare variants in oncogenically associated genes that co-segregate with lung cancer provides specific mutations for future work and improves our understanding of the inheritance and pathogenesis of lung cancer. This work is ongoing.

PrgmNr 2347 - Immune-related intra-tumoral gene expression and survival by *BRCA1/2* mutation status in high grade serous ovarian cancer cases: an Ovarian Tumor Tissue Analysis (OTTA) study

View session detail

Author Block: **R. Vierkant**[1], B. M. McCauley[1], J. M. Cunningham[2], M. S. Anglesio[3], D. D. Bowtell[4], A. DeFazio[5], M. J. Garcia[6], J. Gronwald[7], B. Y. Karlan[8], N. S. Meagher[9], P. D. P. Pharoah[10], Ovarian Tumor Tissue Analysis Consortium, J. A. Doherty[11], A. Talhouk[12], S. J. Winham[1], S. J. Ramus[9], E. L. Goode[1]; [1]Quantitative Hlth.Sci., Mayo Clinic, Rochester, MN, [2]Mayo Clinic & Fndn, Rochester, MN, [3]Peter MacCallum Cancer Ctr., Vancouver, BC, Canada, [4]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [5]The Westmead Inst. for Med. Res., The Univ. of Sydney, Westmead Hosp., Sydney, Australia, [6]Centro Natl. de Investigaciones OncolÃ³gicas, Madrid, Spain, [7]Pomeranian Med. Univ.,, Szczecin, Poland, [8]Univ. of California at Los Angeles, Los Angeles, CA, [9]Univ. of NSW Sydney, Sydney, Australia, [10]Univ Cambridge, Cambridge, United Kingdom, [11]Univ. of Utah, Salt Lake City, UT, [12]Univ. of British Columbia, Vancouver, BC, Canada

Disclosure Block: **R. Vierkant:** None.

Background: The immune system appears to play an important role in determining outcome in women with high-grade serous ovarian cancer (HGSOC) and *BRCA1/2* mutated HGSOC tumors are prime candidates for immunotherapy combined with PARP inhibitors. However, little is known about the expression profile of immune-related genes in tumor tissue and the effect on prognosis for patients with and without *BRCA1/2* mutations. We investigated associations between intra-tumoral expression in nine immune-related genes and survival by *BRCA1/2* mutation status in a multi-site study of women with HGSOC from the Ovarian Tumor Tissue Analysis (OTTA) consortium.

Methods: Nine genes were examined in FFPE tissue containing >70% tumor using NanoString nCounter technology: *CD3E, CD8A, TLR4, MYD88, PD-1, PD-L1, CD68, TAP1 and CD7*4. Gene-specific individual expression values were standardized by subtracting the sample mean and dividing by the sample standard deviation. Associations of expression with 10-year survival, overall and by germline pathogenic *BRCA1* and *BRCA2* mutation status, were modeled using Cox proportional hazards regression adjusted for age, tumor stage and study site. Tests for interaction assessed differential associations between expression and survival by mutation status.

Results: Of 3,055 HGSOC cases included, 2,109 (69%) died within 10 years. Overall, higher expression of *CD3E* (HR 0.94, 95% CI 0.90-0.98); *CD8A* (HR 0.96, CI 0.92-1.00); *PD-1* (HR 0.92, CI 0.88-0.96); *PD-L1* (HR 0.90, CI 0.86-0.94) and *TAP1* (HR 0.85, CI 0.82-0.89) were associated with decreased risk of death. Subset analyses of 640 *BRCA1/2* mutation-negative individuals and 127 *BRCA1* mutation carriers revealed null or protective associations for each gene. In contrast, in 69 *BRCA2* mutation carriers, high expression was associated with increased risk of death for *PD-L1* (HR 1.52, CI 1.04-2.24, p-interaction=0.02) and *CD8A* (HR 1.35, CI 0.97-1.88, p-interaction 0.07). High *TLR4* expression was also associated with risk of death in BRCA2 carriers (HR 1.43, CI 0.98-2.08), but the interaction between TLR4 and mutation status was not statistically significant (p=0.40).

Conclusion: Herein, *BRCA2*-mutated HGSOC cases had different immune-related prognostic associations than those who were mutation-negative or had *BRCA1* mutations. This information may help inform future mutation-specific treatment modalities.

PrgmNr 2348 - Malignant pleural mesothelioma: germline variants may steer tailored treatment

View session detail

Author Block: **M. La Vecchia**[1], M. Sculco[1], A. Aspesi[1], M. G. Clavenna[1], G. Pinton[2], E. Casalone[3], A. Allione[3], F. Grosso[4], R. Libener[5], A. Muzio[6], O. Rena[7], G. Baietto[7], S. Parini[7], R. Boldorini[8], E. Migliore[9], D. Mirabelli[9,10], L. Moro[2], C. Magnani[11], D. Ferrante[11], G. Matullo[3,10,12], I. Dianzani[1,10]; [1]Dept. of Hlth.Sci., UniversitÃ del Piemonte Orientale, Novara, Italy, [2]Dept. of Pharmaceutical Sci., UniversitÃ del Piemonte Orientale, Novara, Italy, [3]Dept. of Med. Sci., UniversitÃ di Torino, Torino, Italy, [4]Mesothelioma Unit, AO SS. Antonio e Biagio e Cesare Arrigo, Alessandria, Italy, [5]Dept. of Integrated activities Res. and Innovation, AO SS. Antonio e Biagio e Cesare Arrigo, Alessandria, Italy, [6]Div. of Med. Oncology, Ospedale Santo Spirito, Casale Monferrato (Alessandria), Italy, [7]Thoracic Surgery Unit, AOU Maggiore della CaritÃ , Novara, Italy, [8]Dept. of Hlth.Sci., Section of Pathological Anatomy, UniversitÃ del Piemonte Orientale, Novara, Italy, [9]Unit of Cancer Epidemiology, CPO-Piemonte and UniversitÃ di Torino, Torino, Italy, [10]InterDept.al Ctr. for Studies on Asbestos and other Toxic Particulates â□□G. Scansettiâ□□, UniversitÃ di Torino, Torino, Italy, [11]Dept. of Translational Med., Unit of Med. Statistics, UniversitÃ del Piemonte Orientale and Cancer Epidemiology, CPO Piemonte, Novara, Italy, [12]Med. Genetics Unit, AOU CittÃ della Salute e della Scienza di Torino, Torino, Italy

Disclosure Block: **M. La Vecchia:** None.

Notwithstanding the direct correlation between the level of asbestos exposure and the risk of malignant pleural mesothelioma (MPM), only a small proportion of individuals with a high asbestos exposure develop MPM. This observation, reports of familial MPM, and identification of germline pathogenic variants (PVs) in *BAP1* or other DNA repair genes in approximately 10% of MPM patients suggest the occurrence of inherited predispositions.

We aimed to search for new predisposing genes, assess the prevalence of PVs in DNA repair genes in MPM patients, and evaluate whether these patients could be sensitive to tailored treatments.

A total of 206 MPM patients (93 from Betti *et al.* 2017, 113 new) were screened by targeted-NGS for germline PVs in cancer-predisposing genes. Six further patients with family history of mesothelioma were analyzed by Sanger sequencing of *BAP1* and *CDKN2A*. Life-long cumulative asbestos exposure was quantified for 203/212 patients.

We identified 18 PVs in 17/206 patients (8.25%), most of them (14 PVs in 13 patients) were found in genes involved in the DNA repair pathway (i.e. *ATM, BRCA1, BRCA2, BRIP1, CHEK2, FANCC, FANCF, FANCI, PALB2, PMS1, SLX4, XPC*). PVs in *BAP1* and *CDKN2A* were identified in five and one patients with family history of mesothelioma, respectively. Carriers of PVs in DNA repair genes (18/212 patients) showed a statistically significant lower asbestos exposure than non-mutated patients (p=0.0001). These data suggest that patients with germline mutations in DNA repair genes are less proficient at repairing the DNA damage induced by asbestos and show increased susceptibility to asbestos-induced MPM.

According to the concept of BRCAness, MPM patients with germline PVs in DNA repair genes may respond to drugs that induce synthetic lethality, a mechanism by which two, otherwise non-lethal defects, become lethal when they are both present in a cell. Thus, we created a 3D-MPM cell model that had a defect in ATM, the master regulator of DNA repair, using a siRNA targeting *ATM* or a specific inhibitor (KU55933). Spheroids obtained using this model showed apoptosis induction when treated with an EZH2 inhibitor (tazemetostat) (p â□¤ 0.05). These data suggest that the subset of patients carrying PVs in DNA repair genes may benefit from this treatment that induces synthetic lethality.

PrgmNr 2349 - On cross-ancestry cancer polygenic risk scores

View session detail

Author Block: **L. G. Fritsche**[1], Y. Ma[1], D. Zhang[1], M. Salvatore[1], S. Lee[2], X. Zhou[2], B. Mukherjee[2]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ MICHIGAN, Ann Arbor, MI
Disclosure Block: **L.G. Fritsche:** None.
Polygenic risk scores (PRS) can provide useful information for personalized risk stratification and disease risk assessment, especially when combined with non-genetic risk factors. However, their construction depends on the availability of summary statistics from genome-wide association studies (GWAS) independent from the target sample. For best compatibility, it was reported that GWAS and the target sample should match in terms of ancestries. Yet, GWAS, especially in the field of cancer, often lack diversity and are predominated by European ancestry. This bias is a limiting factor in PRS research.

By using electronic health records and genetic data from the UK Biobank, we contrast the utility of breast and prostate cancer PRS derived from external European-ancestry-based GWAS across African, East Asian, European, and South Asian ancestry groups.

We highlight differences in the PRS distributions of these groups that are amplified when PRS methods condense hundreds of thousands of variants into a single score. While European-GWAS-derived PRS were not directly transferrable across ancestries on an absolute scale, we establish their predictive potential when considering them separately within each group. For example, the top 10% of the breast cancer PRS distributions within each ancestry group each revealed significant enrichments of breast cancer cases compared to the bottom 90% (odds ratio of 2.81 [95%CI: 2.69,2.93] in European, 2.88 [1.85, 4.48] in African, 2.60 [1.25, 5.40] in East Asian, and 2.33 [1.55, 3.51] in South Asian individuals).

Our findings highlight a compromise solution for PRS research to compensate for the lack of diversity in well-powered European GWAS efforts while recruitment of diverse participants in the field catches up.

PrgmNr 2350 - Rates of mutations in DNA repair genes are low in localized prostate cancer patients, particularly among non-White individuals

View session detail

Author Block: **K. N. Maxwell**[1], R. Hausler[1], A. N. Le[1], G. Kelly[1], J. Powers[2], J. Ding[1], E. Feld[1], H. Desai[1], C. Morrison[1], A. Doucette[1], P. Gabriel[1], M. Jones[3], R. Judy[1], J. Weaver[1], R. Kember[2], S. M. Damrauer[1], D. J. Rader[4], S. M. Domchek[1], V. Narayan[1], L. E. Schwartz[1], D. J. Lee[1]; [1]Univ Pennsylvania, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Regeneron Genetics Ctr., Tarrytown, NY, [4]Univ of Pennsylvania, Philadelphia, PA

Disclosure Block: **K.N. Maxwell:** None.

Background: The identification of germline mutations in DNA repair genes has significant implications for the personalized treatment of individuals with prostate cancer (PrCa). **Objective:** To determine DNA repair genes associated with localized PrCa in a diverse academic biobank cohort and to determine genetic testing burden. **Design, Setting and Participants:** Cross-sectional study of 2,391 localized PrCa patients. **Outcome Measurements and Statistical Analysis:** Genetic ancestry and mutation rates (excluding somatic interference) in 17 DNA repair genes was determined in 1,588 localized PrCa patients and 3,273 cancer-free male controls. Burden testing within individuals of genetically determined European ancestry (EUR) and African ancestry (AFR) was performed between biobank PrCa cases and cancer-free biobank and gnomAD controls. **Results and Limitations:** AFR individuals with localized PrCa had lower DNA repair gene mutation rates than EUR individuals (1.4% vs 3.7%, p=0.04). Mutation rates in localized PrCa patients was similar to biobank and gnomAD controls (EUR: 4.0% vs 2.8%, p=0.15; vs 3.1%, p=0.04; AFR: 1.4% vs 1.8%, p=0.79; vs 2.1%, p=0.52). Gene-based rare variant association testing revealed no gene had significantly enriched mutation rates between biobank PrCa cases and controls in either AFR or EUR. *BRCA2* mutations were significantly enriched compared to gnomAD controls in EUR (1.0% vs 0.3%, p=0.03). 19% and 13% met high-risk and very high-risk criteria; of those, 3.3% and 7.5% had any germline genetic mutation and 1.1% and 2.7% had a *BRCA2* mutation, respectively. Limitations of this study include analysis of a relatively small, single-institution cohort. **Conclusions:** DNA repair gene germline mutation rates are low in an academic biobank cohort of localized PrCa patients, particularly among individuals of AFR genetic ancestry. Only *BRCA2* mutations are enriched compared to non-cancer controls. Mutation rates exceed 5% in very-high risk localized PrCa patients. These findings highlight the importance of germline genetic testing for patients with localized PrCa.

PrgmNr 2351 - The IMPACT study Lynch Syndrome cohort - results after first round of screening

View session detail

Author Block: **E. K. Bancroft**[1], E. C. Page[2], The IMPACT Study Steering Committee and Collaborators, R. Eeles[3]; [1]Royal Marsden Hosp, Sutton, United Kingdom, [2]Inst. of Cancer Res., Sutton, United Kingdom, [3]The Inst. of Cancer Res., Sutton, Surrey, United Kingdom
Disclosure Block: **E.K. Bancroft:** None.
Introduction: Pathogenic variants in the mis-match repair genes (MMR) have been reported to increase the risk of early-onset aggressive prostate cancer (PrCa). The IMPACT study is prospectively evaluating Prostate Specific Antigen (PSA) screening in men with germline MMR pathogenic variants. Here we report the utility of PSA screening, PrCa incidence and tumour characteristics after the first screening round among men with and without these germline pathogenic variants.
Methods: Healthy men aged 40-69 years with germline pathogenic variants *in MLH1, MSH2, and MSH6 genes* and male controls testing negative for a familial pathogenic variant in these genes, were recruited over 90 months from 34 centres in 8 countries. Participants underwent PSA screening and if PSA >3.0ng/ml, were offered prostate biopsy. Screening outcome was assessed related to mutation status.
Results: 204 *MLH1* carriers, 199 *MLH1* non-carriers, 305 *MSH2* carriers, 210 *MSH2* non-carriers, 135 *MSH6* carriers, 177 *MSH6* non-carriers were recruited. Within the first round of screening 56 men had PSA>3.0ng/ml, 35 biopsies were performed, and 18 PrCas diagnosed (13 *MSH2* carriers, 1 *MSH2* non-carrier; 4 *MSH6* carriers). Zero cancers were detected in the *MSH6* non-carriers, and *MLH1* carrier and non-carrier cohort. Cancer incidence was higher in *MSH2* carriers than non carriers (4.3% vs 0.5%; p=0.01) and *MSH6* carriers than non-carriers respectively (3% vs 0%; p=0.04). *MSH2* carriers were diagnosed younger (60 vs 66 years) and more likely to have clinically-significant disease than non-carriers (85% vs 0%). *MSH6* carriers were diagnosed at 64 years (median age) and 75% had clinically significant disease.
Discussion: After the first screening round, carriers of *MSH2* and *MSH6* pathogenic variants had a higher incidence of PrCa, were diagnosed at a younger age and had more clinically significant disease compared with non-carriers. These findings support the use of targeted PSA screening in these men to identify clinically significant disease. Future study screening rounds will help determine the optimal frequency of PSA testing, the utility of PSA screening in *MLH1* carriers and provide further data on the value of annual screening in *MSH2* and *MSH6* carriers.

PrgmNr 2352 - Wood stove use interacts with genetic variants to alter risk of gastric cancer in a high altitude Central American population

View session detail

Author Block: **A. K. Miller**[1], S. B. Rifkin[2], R. L. Dominguez[3], E. Martinez[4], D. Norwood[3], E. E. Montalvan[4], T. Waterboer[5], M. Beasley[6], D. R. Morgan[6], S. M. Williams[1]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Univ. of Michigan, Ann Arbor, MI, [3]Hosp. de Occidente, Copan, Honduras, [4]Hosp. Evangelico, Siguatepeque, Honduras, [5]German Cancer Res. Ctr., Heidelberg, Germany, [6]Univ. of Alabama at Birmingham, Birmingham, AL

Disclosure Block: **A.K. Miller:** None.

Gastric adenocarcinoma is the leading global cause of infection-related cancer mortality and the overall third leading cause of cancer death. In Central America, a disproportionate burden of disease is concentrated in the mountainous regions along the Pacific littoral, a phenomenon often described as the â□□Latin America Altitude Enigma.â□□ Wood stoves (â□□fogonsâ□□) are the predominant method for the heating of homes and cooking, and food preparation with these stoves may be linked to aero-digestive cancers, mediated by ingested and inhaled carcinogens. To assess whether wood stove use is associated with and/or modifies the genetics of gastric cancer, we conducted a population-based, case-control study (814 incident cases, 1,049 controls) in rural western Honduras, a high incidence region with a homogeneous diet, and endemic *H. pylori* infection, predominantly with the high-risk cagA genotype. We investigated the association between gastric cancer and physical and environmental factors including age, sex, *H. pylori* CagA serostatus, wood stove use, 15 variants from 6 heterocyclic amine (HCA) metabolizing genes, and interactions between wood stove use and the HCA-metabolizing gene variants. Age, sex, wood stove use, and *H. pylori* CagA infection were independently associated with gastric cancer in univariate analyses, as well as two SNPs in the gene *CYP1B1* (rs1800440 and rs1056836). Rs1800440 remained significant in models adjusting for combinations of variables age, sex, CagA serostatus, woodstove use, and rs1056836. When the interaction between rs1800440 and wood stove use was added to any model, rs1800440 as an independent predictor of gastric cancer was no longer significant, but the effect size of wood stove use increased by almost 2-fold when the interaction was included in the model. Lifetime wood stove use is associated with gastric cancer, independent of *H. pylori* CagA serostatus, in the high incidence regions of mountainous Central America where *H. pylori* infection is endemic. However, the impact of this exposure on disease risk is dependent on *CYP1B1* genotype, indicating the gene-environment interaction plays a significant role in the association with gastric cancer. Therefore, environment and genetic risk factors should be considered together in assessing risk and planning surveillance and prevention.

PrgmNr 2353 - ACSL VI:A Novel Variation Results in Structural Modification and Multiple Neoplasia in a 46-year-old Female

View session detail

Author Block: **I. Castillo**[1], E. E. Freire[1], V. Romero[1], K. Hosomichi[2], A. Tajima[2], B. Arias[3], C. Reyes[4]; [1]Univ. San Francisco de Quito, Quito, Ecuador, [2]Kanazawa Univ., Kanazawa, Japan, [3]Onelabt S.A, Ballenita, Ecuador, [4]Hosp. Eugenio Espejo, Quito, Ecuador

Disclosure Block: **I. Castillo:** None.

Multiple non-related neoplasia in one patient is a rare event and suggest an inherited pathological variant and a further loss of the normal allele. We analyzed a 46-year-old female patient diagnosed with frontotemporal astrocytoma (2010), papillary thyroid cancer (2013), and ductal breast cancer in 2017 and no family history. The general population risk for developing these three neoplasia within one lifespan is 0.000000000675 compared to 0.00000342 of winning the lottery.

We hypothesized a shared affected gene responsible for the conditions, not frequently associated with cancer. We performed whole-exome sequencing (WES) identifying a novel missense variant in the ACSL VI gene and no pathological variations in common neoplasia-associated genes. ACSL VI catalyzes long-chain fatty acids mainly in the brain (mainly expressed in the astrocytes), and in minor quantities in the breasts, thyroid, erythrocytes, leucocytes, and other tissues. Notably, the ACSL gene family has been shown to be implicated in the development and prognosis of different types of cancer.

Next, we modeled the mutant ACSL VI confirming the damaging effect in the secondary and tertiary structures. The details found in the ACSL VI gene allow us to interpret that the mutant protein (patient) presents a change in the secondary and tertiary structure important for anchoring to the membranes of the endoplasmic reticulum, mitochondria, and peroxisomes. As it is an evolutionarily conserved region, its function is essential for correct cell development and metabolism, making it an ideal candidate in the hypothesis that this is the gene associated with the patient's cancers.

Finally, our research suggests the importance of the study of non-conventional cancer-associated genes involved in lipidic metabolic processes. This can open the doors for the development of new personalized therapeutic approaches targeting the affected gene and potential metabolic pathways involved in neoplastic processes.

PrgmNr 2355 - Functional study of the 2q22 renal cancer susceptibility locus

View session detail

Author Block: **A. G. Souza**[1], L. Jessop[2], T. A. Myers[2], M. Machiela[2], K. Brown[2], S. Chanock[2], L. Colli[1,2]; [1]Dept. of Med. Imaging, Hematology, and Oncology, RibeirÃ£o Preto Med. Sch. - Univ. of SÃ£o Paulo, RibeirÃ£o Preto, Brazil, [2]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, Dept. Hlth.and Human Services, Bethesda, MD

Disclosure Block: **A.G. Souza:** None.

Renal cancer (RCC) tumorigenesis develops partly due to series of somatic genetic events such as mutations, deletions and epigenetic alterations in tumor suppressor genes. Genome Wide Association Study (GWAS) have identified 13 different genomic regions associated with the risk of developing RCC. We report on analyses of the 2q22 region, whose biological mechanisms related to the development of RCC are not well understood. Thus, the aim of this study was to identify the role of functional SNPs located in the 2q22 region based on the assessment of putative functional SNPs identified and prioritized using MPRA, ATAC-Seq, ChIP-Seq, Capture-HiC, eQTL, LDlink data and the validation of the SNP by electrophoretic mobility shift assays (EMSA) assays. Among the 14 SNPs evaluated, we selected the SNP rs72858496 that showed high linkage disequilibrium (R^2 = 0.93; Dâ = 1) with the SNP tag (rs12105918) with a *p-value* of 2.9 x 10^{-9}. We have shown that this SNP has a chromosomal loop that allows physical interaction with the promoter region of the *ZEB2* gene. EMSA analyses of this SNP revealed that the DNA fragments produced a specific shift with nuclear extracts from renal cancer cell line (RCC4). Thus, our initial results point to rs72858496 as a potential functional SNP in the 2q22 region, altering *ZEB2* expression and consequently contributing to the progression of renal cancer. Additional functional studies are now underway to understand the molecular and oncological mechanisms linked to the 2q22 region that may contribute to a better understanding of RCC, which could also impact the treatment and prognosis of patients with this disease.

PrgmNr 2356 - Inflammatory and oncological processes contributing to the increased susceptibility to Barrettâ□□s esophagus in adults born with esophageal atresia

View session detail

Author Block: **E. Brosens**[1], C. A. ten Kate[1], B. de Graaf[1], M. Doukas[1], A. Koivusalo[2], H. Nastiti[1], R. van der Helm[1], T. Brands[1], H. IJsselstijn[1], Y. van Bever[1], A. de Klein[1], R. M. H. Wijnen[1], M. C. Spaander[1], R. M. W. Hofstra[1]; [1]Erasmus Univ. Med. Ctr. â□□ Sophia Childrenâ□□s Hosp., Rotterdam, Netherlands, [2]Univ. of Helsinki, Children's Hosp., Helsinki, Finland, Helsinki, Netherlands
Disclosure Block: **E. Brosens:** None.
Introduction Adults born with esophageal atresia (EA) are more prone to chronic gastroesophageal reflux (GER), which can cause Barrettâ□□s esophagus (BE). The prevalence of BE in EA adults is 4-times higher than in the general population and presents at a younger age (34 versus 60 years). Given the overlap between genes and pathways involved in foregut development and risk loci and pathways for BE, we hypothesized that EA patients have a predisposition to develop BE.
Material and methods Blood and esophageal biopsies were collected during endoscopic surveillance. We compared risk loci and transcriptomes of 19 EA patients with BE (EA/BE); 44 EA patients without BE (EA only); 10 BE patients without EA (BE only) and 730 unaffected controls. Subsequently, we simulated a reflux episode by exposing fibroblasts of 3 EA patients and 3 unaffected controls to acid medium, and compared their response.
Results We found a median polygenic risk score of 3.24 (range 1.39-4.68) in EA/BE patients versus 2.63 (1.85-3.53) in BE only patients. Pathway enrichment analysis revealed differences in retinoic acid signaling and downstream pathways as well as inflammatory, stress response and oncological processes in EA/BE patients relative to BE only patients. In-vitro experiments in fibroblasts confirmed the effect on retinoic acid signaling and immune response in EA patients upon acid exposure.
Conclusion The contribution of risk loci seems elevated, although sample sizes are insufficient to draw definite conclusions. Epithelial tissue homeostasis is more prone to disturbances in EA patients. This intrinsic susceptibility could potentially explain the earlier age of onset of BE in these patients.

PrgmNr 2358 - Utility of altered nuclear expression of the *BAP1* for assessment of pathogenicity of Variants of Unknown Significance

View session detail

Author Block: **P. Hanpude**[1], D. Etia[1], J. B. Massengill[1], F. H. Davidorf[1], L. Byrne[2], C. M. Cebulla[1], M. H. Abdel-Rahman[3]; [1]Ophthalmology, The Ohio State Univ., Columbus, OH, [2]Div. of Human Genetics, The Ohio State Univ., Columbus, OH, [3]The Ohio State Univ., Columbus, OH
Disclosure Block: **P. Hanpude:** None.
Purpose- Germline mutation in the tumor suppressor gene *BAP1* is associated with the hereditary tumor predisposition syndrome, *BAP1*-TPDS (OMIM 614327), which is associated with predisposition mainly to four cancers: uveal melanoma, mesothelioma, cutaneous melanoma, and renal cell carcinoma. A growing number of variants are being deposited in ClinVar, including 706 variants of uncertain significance (VUS). It has been reported that variants with loss of deubiquitinase function of BAP1 are associated with loss of nuclear BAP1 (nBAP1) localization. We hypothesized that detection of loss of nBAP1 would help to characterize the pathogenesis of VUSs.
Methods- Flag-tagged full-length *BAP1* was sub-cloned in p3xFLAG-CMV-10 vector. Mutations were generated by site-directed mutagenesis and confirmed by Sanger sequencing. We transiently overexpressed empty vector, wild type and mutant BAP1 in NCI-H226 cell line with no endogenous BAP1 and BAP1 knock out (KO) SNU449 cells. Localization of all the mutant proteins was explored by immunofluorescence (IF) staining with anti-BAP1 and anti-Flag antibodies. Samples were analyzed by a confocal microscope.
Results- We selected four VUSs (L14H, N78S, Q85H and W202R), three truncating variants (Q267*, Q393* and Q573*), a benign variant; G41S, catalytic inactive pathogenic variantC91A and an NLS (Nuclear Localization Signal) mutant. *BAP1* deletion and loss of expression in NCIH226 and SNU449 BAP1 KO cells were confirmed by sequencing and western blot, respectively. Wild type BAP1 as well as the benign G41S variant showed both nuclear and cytoplasmic expression. NLS and truncating variants showed complete loss of nuclear expression. Catalytic inactive C91A variant showed perinuclear accumulation along with decrease nuclear expression confirming the importance of auto-deubiquitination for the nuclear localization of BAP1. Among the tested VUSs, W202R, Q85H and N78S showed loss of BAP1 nuclear localization. Interestingly, the VUS L14H showed partial loss of nuclear localization. Segregation studies of the W202R variant suggest that it is a likely pathogenic variant.
Conclusions- Our results suggest that BAP1 nuclear localization might be used to assess the pathogenesis of *BAP1* missense VUSs. Validation studies in a larger cohort of well-characterized missense *BAP1* VUSs is warranted.
This work was supported by funds from the R21CA219884 grant from the National Cancer Institute

PrgmNr 2359 - Chromatin accessibility mapping in T-cell Prolymphocytic Leukemia

View session detail

Author Block: **H. Yan**[1], S. Tian[1], H. Jin-Lee[1], H. Zhang[2], P. Hampel[2], E. Klee[1], W. Ding[2]; [1]Div. of Computational Biology, Mayo Clinic, Rochester, MN, [2]Div. of Hematology, Mayo Clinic, Rochester, MN
Disclosure Block: **H. Yan:** None.
T cell prolymphocytic leukemia (T-PLL) is a rare disease with a median survival of

PrgmNr 2360 - DNA Methylation Profiles of Ovarian Clear Cell Carcinoma

View session detail

Author Block: **J. M. Cunningham**[1], S. J. Winham[2], C. Wang[3], B. Weiglt[4], Z. Fu[5], S. Armasu[3], B. McCauley[3], A. Brand[6], Y-E. Chiew[6], E. Elishaev[7], C. Gourley[8], C. Kennedy[6], A. Laslavic[5], J. Lester[9], A. Pizkorz[8], M. Sekowsky[10], A. DeFazio[6], B. Karlan[9], J. Brenton[10], M. Churchman[8], R. Drapkin[11], K. Elias[12], D. Huntsman[13], J. Konner[4], K. Lawrenson[14], E. Papaemmanuil[4], F. Modugno[5], K. Bolton[15], E. Goode[1]; [1]Mayo Clinic & Fndn, Rochester, MN, [2]Mayo Clinic, Rochester, MN, [3]Quantitative Hlth.Sci., Mayo Clinic, Rochester, MN, [4]Mem. Sloan Kettering Cancer Ctr., New York, NY, [5]Univ. of Pittsburgh Sch. of Publ. Hlth., Pittsburgh, PA, [6]Westmead Hosp., Sydney, Australia, [7]elisex@upmc.edu, Pittsburgh, PA, [8]Univ. of Edinburgh, Edinburgh, United Kingdom, [9]Univ. of California at Los Angeles, Los Angeles, CA, [10]Univ. of Cambridge, Cambridge, United Kingdom, [11]Univ. of Pennsylvania, Philadelphia, PA, [12]Brigham and Women's Hosp., Boston, MA, [13]Univ. of British Columbia, Vancouver, BC, Canada, [14]Cedars-Sinai Med. Ctr., Los Angeles, CA, [15]Univ. of St Louis, St Louis, MO

Disclosure Block: **J.M. Cunningham:** None.

Background: Ovarian clear cell carcinoma (OCCC) is rare and tends to respond poorly to standard platinum-based chemotherapies, the mainstay of ovarian cancer. Tumor methylation profiles differ between OCCC and other histotypes; but studies have been too limited in sample size to evaluate variation within the OCCC histotype. Here, we sought to define the role of DNA methylation and prognostic subclassification of a large series of OCCC.

Methods: To better understand the role of DNA methylation in the biological and prognostic subclassification of OCCC, we used Illumina Infinium MethylationEPIC and HumanMethylation450k Beadchips to interrogate methylation in tumor DNA from 271 fresh frozen OCCC tumors from ten study sites.

Results: Non-smooth non-negative matrix factorization clustering revealed two approximately equally sized clusters defined by the most variable 2,437 CpGs. Cluster membership was associated with several clinical features: compared to Cluster 2 (N=137), Cluster 1 OCCCs (N=134) presented at a more advanced stage, were less likely to be of Asian ancestry, and tended to have poorer outcomes including macroscopic residual disease following primary debulking surgery (p-values Conclusion: This work serves as a foundation for integrative analyses to better understand the complex biology of OCCC in an effort to facilitate development of targeted therapeutics.

PrgmNr 2361 - From genetics to immune system: the dynamic crosstalk in a wide cohort of completely resected Gastrointestinal Stromal Tumors (GISTs)

View session detail

Author Block: **L. Incorvaia**[1], D. Fanale[2], C. Brando[2], L. Algeri[2], A. Dimino[2], R. Scalia[2], N. Barraco[2], M. Bono[2], E. Pedone[2], D. Cancelliere[2], A. Fiorino[2], A. Galvano[2], V. Gristina[2], A. Perez[2], L. Corsini[2], A. Pivetti[2], G. Badalamenti[2], A. Russo[2], V. Bazan[1]; [1]Dept. of Biomedicine, NeuroSci. and Advanced Diagnostics (Bi.N.D.), Section of Med. Oncology, Univ. of Palermo, Palermo, Italy, [2]Dept. of Surgical, Oncological and Oral Sci., Section of Med. Oncology, Univ. of Palermo, Palermo, Italy
Disclosure Block: **L. Incorvaia:** None.
The exon 11 mutations are the most frequent *KIT* mutations, but represent a heterogeneous subgroup in terms of biological and clinical behavior. The exact pathogenic variant (PV) type and codon location, and other biological factors, could affect the Recurrence-Free Survival and the development of an organ-selective pattern of tumor metastasis to relapse. 116 GIST patients completely resected were included in the study between January 2005 and September 2020. The association between Exon 11 PV type with RFS was evaluated. In relapsed patients, metastatic sites were described, Neutrophil-Lymphocyte Ratio (NLR) was calculated, and plasma PD-1, PD-L1, BTN3A1 and BTN2A1 levels have been measured using homemade ELISA assays not yet commercially available. Deletions (del) of the codons 557/558 showed more aggressive tumor behavior and a higher risk of recurrence compared to other exon 11 del, or duplication/insertion/SNV (7-year RFS: 48.6% vs 73.1% vs 87.9%; pKIT p.W557_K558del); when 557 and 558 deletions were analyzed separately, only 22.2% showed a tumor relapse, miming the prognostic behavior of tumor carrying del outside 557/558 position. In contrast to previous findings, in relapsed patients with 557/558 del, the peritoneum is the most frequent metastatic site (72.2%). Using thresholds by ROC analysis and multivariate analysis, we found that the patients with a tumor harboring del outside 557/558 and prevalent peritoneal metastasis, had NLR significantly lower compared to the relapsed patients with other exon 11 PVs (median 2.3 vs 3.1 G/L; p=0.029), and lower baseline levels of plasma PD-1 (>7.9 ng/mL), PD-L1 (

PrgmNr 2362 - Genome Variation Analysis on Accelerated Framework

View session detail

Author Block: **P. Vats**[1], E. Crowgey[2], K. Franke[2], G. Burnett[1], A. Sethia[1], T. Harkins[1], T. Druley[3,4]; [1]Nvidia, Santa clara, CA, [2]Nemours Alfred I duPont Hosp. for Children, Wilmington, DE, [3]Washington Univ. Sch. of Med., St Louis, MO, [4]ArcherDX / Invitae Corp., San Francisco, CA
Disclosure Block: **P. Vats:** Salary/Employment; Nvidia.
In recent years, molecular profiling of cancer patients using next generation sequencing (NGS) has become a crucial clinical tool for cancer care, via the ability to detect clinically actionable variants including single nucleotide variants (SNVs), small InDels, copy number variants (CNVs) and structural variants(SVs). This collection of variants and mutations plays an important role in cancer diagnosis and therapeutic selection. To further improve the effectiveness of NGS data in cancer care, the integration of multi-omics datafrom whole genome, exome, transcriptome, methylome, and clinical data, this provides a more comprehensive molecular profile of the cancer genome. Currently, the major bottlenecks appear to be the processing and integration of multi-omic data, often taking weeks to months to complete, which will only grow as longitudinal multi-omic assessments for the same patient become the clinical standard. By using an accelerated framework, it is possible to iterate the data analysis to reduce the complexity of this computational puzzle while simultaneously identifying those clinically actionable signatures to improve cancer therapy outcomes.
We implemented an accelerated framework that provides 15-50x faster run times for a series of somatic callers including Mutect2, Somatic Sniper, MuSE, and VarScan algorithms, allowing more variants to be identified from a data set. First, we assessed the performance of the individual baseline variant calling algorithms on the SEQC-II 50x WGS data set. For somatic SNV detection, Strelka2 had the highest F1 score, and recall (0.9435, 0.9425) while MuSE had the highest precision (0.9966). For somatic InDel detection, Mutect2 and Strelka2 performed similarly with Mutect2 higher recall, but Strelka2 had higher F1 score and precision.
Here, we apply the Clara Parabricks accelerated framework to 458 samples from leukemia patients treated on the same phase III randomized clinical trial. Analyses focused on variant calling, copy number analysis and fusion calling from genome, transcriptome and targeted DNA or RNA deep sequencing panels, providing the basis for a multi-omics approach to identify those genetic signatures that correlate with disease.

PrgmNr 2364 - Germline variants associate with the presence of somatic *TP53* and *PIK3CA* mutations in breast tumors

View session detail

Author Block: **A. E. Toland**, N. Pollock, J. Ramroop, P. Stevens, J. McElroy; The Ohio State Univ., Columbus, OH
Disclosure Block: **A.E. Toland:** None.
Somatic mutations in *TP53* and *PIK3CA* in breast tumors have unequal prevalence among racial and ethnic groups. Some of the factors leading to somatic mutations in specific genes can be attributed to socioeconomic and environmental exposures, but studies by our group and others suggest that inherited genetic variants also impact selection of mutations in specific genes. To test the hypothesis that germline variants could contribute to selection of somatic *TP53* and *PIK3CA* mutations in breast tumor development, we performed a genome-wide association study of ~776,800 single nucleotide variants in 2850 women with breast cancer using existing germline genotypes and somatic tumor mutation data from The Cancer Genome Atlas, METABRIC, and the Welcome Trust. For initial discovery studies, only data from non-Hispanic whites, defined by principle component analyses, were used. Five SNVs associated with *TP53* somatic mutations were identified with p-values -6 and 34 with p-values -5. Some variants associated with the presence of *TP53* mutation of any type; others were exclusively associated with loss-of-function or gain-of-function *TP53* mutations. Some significant variants were previously associated with estrogen receptor negative breast cancer in GWAS. This includes rs9397437 that maps to the estrogen receptor 1 locus and was shown in previous studies to affect expression of *ESR1*, *RMND1* and *CCDC170*. We identified 44 SNVs associated with *PIK3CA* mutations with p-values -5. Some variants were associated with the presence of any *PIK3CA* somatic mutation and others only with a hotspot mutation (e.g. H1047R, E542K or E545K). These studies suggest that germline variants can shape a cellular environment which promotes selection of somatic mutations in specific genes and provide insight into which genes/pathways may be important for the cellular context that supports tumors with these specific mutations. Interestingly, some of these variants map to previous GWAS loci for breast cancer risk, specifically triple negative breast cancer risk, suggesting that there may be overlap of variants associated with risk of tumor subtypes and specific mutations. Ongoing studies are evaluating these variants for association with *TP53* and *PIK3CA* somatic mutations using more diverse study populations.

PrgmNr 2365 - Identify non-mutational TP53 loss of function in human cancers

View session detail

Author Block: **L. WANG**[1], Q. Li[2], Y. Zhang[2], L. Ma[2], Z. Zhang[2]; [1]Mayo Clinic, ROCHESTE, MN, [2]Beijing Inst. of Genomics, Beijing, China
Disclosure Block: **L. Wang:** None.
Background: A gene with an intact DNA sequence can compromise its function by epigenomic, transcriptomic, and proteomic level dysregulations. Such non-mutational inactivations are prevalent in cancers but, due to the heterogenetic causes, cannot be detected by the DNA-sequencing, immunohistochemical staining, or any other single assay alone. Therefore, they would become a significant impediment for molecular diagnostics, clinical management, and treatment selection for cancer patients. **Approach**: We hypothesized that when a transcription factor (TF) is functionally impaired, either due to genetic or non-genetic causes, the expression of downstream target genes would be significantly altered and that such expression alteration, in turn, can be used to predict the functional status of the TF. Here we used p53 as an example; we first define the p53 target genes through a comprehensive literature review and meta-analysis. Then, we build an SVM model using the composite expression scores of these target genes as features and using the â□□normal tissuesâ□□ (assuming p53â□□s tumor suppressor function is normal in this group) and â□□*TP53* truncating tumor samplesâ□□ (assuming p53â□□s tumor suppressor function is lost in this group) as training datasets. **Results**: Using 5-fold cross-validation, we demonstrated the superior performance of our SVM model (average AUC = 0.995, F1-score = 0.989, recall = 0.992) in TCGA LUNG and BRCA cohorts. When applying our SVM model to *TP53*^{WT} tumor samples, we found 87% of BRCA and 94% of LUNG samples were predicted to be LoF (termed as *TP53*^{WT}-LoF). These *TP53*^{WT}-LoF patients exhibited distinct genomic and clinical characteristics from the other *TP53*^{WT} patients. Specifically, *TP53*^{WT}-LoF patients have significantly higher tumor mutation burden, the fraction of genome with copy number variations, aneuploidy score, and hypoxia score, consistent with p53â□□s function as a central regulator of DNA damage repair and cellular stress response. In addition, *TP53*^{WT}-LoF patients with lung cancer have significantly shortened overall survival compared to those real *TP53*^{WT} patients. Further analyses revealed that *MDM2*/*MDM4* amplifications are significantly enriched in *TP53*^{WT}-LoF patients, partially and mechanistically explained the p53 loss-of-function.

PrgmNr 2366 - Integrated Transcriptomic Profiling with Deep Learning Classification Applied to Prostate Cancer Tissue Images

View session detail

Author Block: **D. Van Booven**[1], V. Sandoval[2], F. Ismael[3], A. BriseÃ±o[2], K. Mirjat[4], O. Kryvenko[5], H. Arora[6]; [1]John P Hussman Inst. for Human Genomics, Univ Miami Miller Sch Med, Miami, FL, [2]Hosp. Valentin Gomez Farias, Univ. de Guadalajara Mexico., Guadalajara, Mexico, [3]Dow Univ. of Hlth.Sci., Karachi, Pakistan, [4]Dept. of Pathology, Shaheed Mohtarma Benazir Bhutto Med. Coll., Karachi, Pakistan, [5]Dept of Pathology, Univ Miami Med. Group, Miami, FL, [6]Dept of Urology, Univ Miami Miller Sch Med, Miami, FL

Disclosure Block: **D. Van Booven:** None.

Introduction : Prostate cancer (PCa) diagnosis starts with PSA level detection, which if above the normal range, the patient is subjected to genomic tests like 4K scores, PCA3 or PHI test. After confirmation, MRI is used to identify potential areas of PCa. The biopsy is extracted by the clinicians, inspected by the pathologist and are further subjected to genomic testing. Typically, a single worst area is selected for the test, leaving a large area from consideration. In case of false outcome, patients are subjected to the same steps making it financially cumbersome for the patients and increasing time/accuracy for treatment. We propose a multi-tiered approach using machine learning to analyze the entire section and integrate gene signatures in real time. This allows us to A) identify/characterize the areas of cancer, B) monitor the areas more precisely enabling us to study the therapy response in a longer run. **Methods :** Prostate slide tissues from 500 individuals in the PRAD study contained in TCGA were downloaded and evaluated. We used multiple deep learning algorithms before selecting the Xception model. We selected 107 random images across all Gleason scores (GS) that were then scored by 3 pathologists as a basis for training images to create the network to automate GS grading. This automation identified cancer regions of interest on test training slides with minimal hyperparameter manipulation. Additionally, we created gene expression profiles by taking RNAseq data and combining patients within a specific GS. To reduce variance, these profiles were then put through a rigorous analysis pipeline to identify and remove variability and potential confounding factors. Finally, these profiles were analyzed for differential expression to obtain genes that are uniquely different at a given Gleason grade. **Results :** Heterogeneity is seen when evaluating all 107 images and requiring all 3 pathologists, the AI as well as TCGA agree with only 25% accuracy. When requiring only 2 of 3 pathologists, the AI, and TCGA to agree this rises to 53% accuracy. However, using more generic classification system including normal tissue, low grade tumor (GS6 or GS7), and high grade tumor (GS8 or GS9) this accuracy is improved to 71%. Finally, integration with the transcriptome classifier this was further enhanced to 81%. **Conclusions :** ML can be applied to determine severity of tissue images. We are hopeful that future modifications will improve the accuracy of our current model and will aid in identifying tumor areas, accessing their severity, and enhance treatment decisions.

PrgmNr 2367 - Integrative genomic analysis reveals new insights on complex mutational landscape in T-cell prolymphocytic leukemia

View session detail

Author Block: **S. Tian**[1], H. Zhang[2], H. Jin-Lee[1], P. Hampel[2], E. Klee[1], H. Yan[1], W. Ding[2]; [1]Div. of Computational Biology, Mayo Clinic, Rochester, MN, [2]Div. of Hematology, Mayo Clinic, Rochester, MN
Disclosure Block: **S. Tian:** None.
T-cell prolymphocytic leukemia (T-PLL) is a rare mature T-cell malignancy with aggressive clinical course and poor prognosis. Cytogenetic, whole-exome and whole-genome analysis have identified major structural variations in T-PLL, including inv(14) (q11q32) and t(14;14)(q11;q32) that activated the oncogenic expression of TCL1A as a hallmark. In addition, recurrent somatic mutations were identified in epigenetic regulators, tumor suppressors and genes from the JAK-STAT signaling pathway. However, T-PLL demonstrates a strong heterogeneity, and our understanding of the molecular events driving T-PLL pathogenesis is mostly from small cohort studies and thus remains incomplete. This study analyzed whole-exome and transcriptome data from a cohort of 14 cases. We identified recurrent mutations previously known in T-PLL, such as ATM (57%), STAT5B (28.6%), and JAK3 (21.5%), along with novel recurrent mutations that may play critical roles in T-PLL. Specifically, the latter include genes in cancer initiation and proliferation (PABPC3/71.4%, HRNR/64.3%, AK2/50%, LRP1/35.7%, and NOTCH2/14.3%), activators of ERK/JNK kinase pathway (MAP3K1/50% and MAP3K4/35.7%), key epigenetic regulators (KDM6B/21.4% and HDAC8/14.3%), and immune surveillance (MAGEC1/64.3% and HLA-DQA2/57.1%). In addition, we identified many recurrent arm-level or large-scale focal events across over 13 chromosomes (15.4-100%). These somatic copy number alterations (SCNAs) were manually inspected through coverage profiles. The most significant arm-level SCNAs were on 8p (9/14) and 8q (7/14). Other significant SCNAs were observed on chr 7 (14/14), followed by chr22 (61.5%), chr12 (53.8%), chr14 (53.8%), chr11 (46.1%) and chr20 (46.1%). The genes targeted by SCNAs are highly associated with cell cycle and chromosome maintenance. Finally, analysis of transcriptome data identified 281 down- and 282 up-regulated genes in T-PLL cases compared to those of healthy controls. As expected, TCL1A and TCL1B were significantly overexpressed (>10 fold). The down-regulated genes are enriched in the cellular defense response, regulation of immune response, chemokine receptor activity, and regulation of T cell proliferation, while up-regulated genes are involved in the regulation of MAPK cascade, MAP kinase activity, and protein kinase binding. In summary, our integrative analyses of genomic and expression data provided new insights into T-PLL complex mutational landscape. Further analysis of the altered key genes and pathways could help elucidate the molecular mechanisms of T-PLL pathogenesis.

PrgmNr 2368 - Investigating associations of common *ATM* missense variants with neoplasms in the UK Biobank exome sequencing data - a conditional analysis

View session detail

Author Block: **X. Jiang**[1], A. O'Neill[1], K. R. Smith[2], Z. Lai[3], K. J. Carss[2], Q. Wang[4], S. Petrovski[1]; [1]Ctr. for Genomics Res., AstraZeneca, Melbourn, United Kingdom, [2]AstraZeneca, Melbourn, United Kingdom, [3]Astra Zeneca, Waltham, MA, [4]Ctr. for Genomics Res., AstraZeneca, Waltham, MA
Disclosure Block: **X. Jiang:** None.
As an essential regulator of DNA damage, Ataxia-telangiectasia mutated (*ATM*) gene has been widely studied in oncology. However, the independent effects of *ATM* missense variants and protein-truncating variants (PTVs) on neoplasms have not been heavily studied in large population biobank settings. Whole-exome sequencing data and the clinical health records of ~400K UK Biobank participants of European ancestry were used in this analysis. We mined genetic associations from variant-level and gene-level phenome-wide association studies and conducted a variant-level conditional association study to test whether the effects of *ATM* missense variants on neoplasms were independent of *ATM* PTV carrier status.

Of 773 distinct protein-coding variants in *ATM*, variant-level association analysis revealed three missense variants significantly (P-8) associated with at least one of 12 phenotypes relating to six different neoplasms. Remarkably, none of the phenotypes significantly associated with these three *ATM* missense variants were among established *ATM* PTV-linked malignancies. Whereas our gene-level PTV collapsing analysis was consistent with established *ATM* PTV literature showing that the aggregated impact of 286 *ATM* PTVs significantly (p-9) associated with 17 malignant neoplasm phenotypes. A subsequent conditional analysis identified that the missense signals were acting independently of the known clinically relevant *ATM* PTVs.

PrgmNr 2369 - JBrowse 2: an extensible open-source platform for modern genome analysis

View session detail

Author Block: **C. Diesh**[1], R. Buels[1], G. Stevens[1], P. Xie[1], H. Elliot[1], T. Martinez[1], C. Bridge[2], E. Yao[3], S. Cain[2], S. Dider[1], L. Stein[4], I. Holmes[1]; [1]Univ. of California at Berkeley, Berkeley, CA, [2]Ontario Inst. of Cancer Res., Toronto, ON, Canada, [3]UC Berkeley, Berkeley, CA, [4]Ontario Institue for Cancer Res., Toronto, ON, Canada
Disclosure Block: **C. Diesh:** None.
Introduction
Genome browsers are useful tools for research in human genetics, able to display and integrate biological information such as long and short read sequencing data, variant calls, and annotations. While traditional linear genome browsers have demonstrated broad utility for many applications, there is an increasing need for visualizing complex structural variants (SVs) beyond the linear view. To meet these challenges, we created JBrowse 2: an extensible open-source platform for visualizing and integrating genomic data.
Results
JBrowse 2 is a flexible platform that provides the foundation for the development of applications and web dashboards that combine novel views and representations extending beyond linear displays in genomic reference coordinates. For example, a flagship JBrowse 2 application is the SV Inspector, which lets users open a list of structural variants in a data table and see the results in a whole-genome circular view. Clicking on a given variant in the table or circular view opens up a linear view that displays the read evidence supporting the SV, even across complex breakpoints like translocations. Dotplot and synteny views are built-in, enabling â□□long read vs referenceâ□□ dotplot visualizations, or alignments of de-novo assembled contigs to the genome. These views are integrated with each other, so that (for example) a long read is just a few clicks away from the corresponding dotplot.
In addition to including these new views, the JBrowse 2 platform is designed from the ground up to enable third-party plugins to add new views, data adapters, and track types, which has facilitated the creation of new plugins that address some very specific use cases. These include our MSA view plugin for viewing multiple sequence alignments, data adapters that download data from UCSC and CIVIC APIs, and the Quantseq plugin for viewing quantitative motif scores as a genome browser track.
JBrowse 2 is available as both a web app or a local desktop app. We also offer re-usable components on NPM, and users of R can programmatically create an instance of JBrowse 2 with the JBrowseR package on CRAN. JBrowse 2 can also generate high quality SVG snapshots from inside the app, and we also created a CLI tool called jb2export to perform automated or bulk exports of JBrowse 2 visualizations. We anticipate that JBrowse 2 will better serve genome scientists with its structural variant visualization capabilities, and will provide the flexibility to adapt to new visualization and analysis challenges as they emerge in the coming years.

PrgmNr 2370 - Oncogenic mutations in the normal human brain

View session detail

Author Block: **J. Ganz**[1,2,3], E. Maury[1,2,3], B. Becerra[1,2], S. Bizzotto[1,3,2], R. N. Doan[1], C. Kenny[1], T. Shin[1,2,3], J. Kim[1,3,2], Z. Zhou[1,3,2], K. L. Ligon[4,5,2,1,3], A. E. Lee[1,3], C. A. Walsh[1,3,2]; [1]Boston Children's Hosp., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Dana-Farber Cancer Inst., Boston, MA, [5]Brigham & Womenâ□□s Hosp., Boston, MA

Disclosure Block: **J. Ganz:** None.

Oncogenic mutations have been found in non-diseased, proliferative tissues such as skin, blood, and esophagus, showing an age-related increase in most cases. However, the prevalence of such mutations in the normal brain, a low-proliferating organ, is unknown and remains challenging due to the abundance of non-proliferating neurons concentrated in the grey matter. Thus, targeting the white matter for sequencing, which is rich in glial cells, and analyzing data from large cohorts would allow for enhanced power to detect oncogenic events in the non-diseased brain. We evaluated genes implicated in brain tumors by deep sequencing in over 418 normal brain samples derived from 110 individuals of varying ages, without tumor diagnosis, and separately analyzing white and grey matter and other brain regions. We detected predicted and reported pathogenic oncogenic variants in cancer driver genes such as *IDH1, PTPN11, NF1,* and *PTEN*. These mutations were predominantly present in the subcortical white matter (p=0.029) and, surprisingly, were less common in older individuals (p=0.025). In addition, we identified the recurrence of glioma driver variants and their enrichment in glial cells. These findings were replicated using 1,640 non-diseased bulk RNA-seq brain samples from the GTEx consortium, confirming the existence of oncogenic variants and depletion of protein disruptive variants with age. The normal brain exhibits enrichment of mutational signatures present in brain tumors (p=0.00018), suggesting that mutational processes of the normal brain drive early oncogenesis. Our study helps understand the origin and early evolution of brain tumors and opens avenues for early interventions and more accurate diagnosis baselines.

PrgmNr 2371 - Paired DNA and RNA genetic testing improves variant detection and classification independently of personal history of cancer

View session detail

Author Block: **S. Campian**, K. Burgess, C. Horton, L. Hoang, R. Pilarski, J. Profato, H. Zimmerman, E. Chao, R. Karam; Ambry Genetics, Aliso Viejo, CA
Disclosure Block: **S. Campian:** Salary/Employment; Ambry Genetics.
Introduction: Recent advances in germline hereditary cancer testing include concurrent analysis of mRNA (RNA genetic testing, or RGT) in conjunction with DNA genetic testing (DGT). RGT can provide an additional line of evidence for variant interpretation, leading to fewer variants of unknown significance (VUS) and increased detection of deep intronic pathogenic/likely pathogenic variants (PVs) that previously were undetected by DNA testing alone.As healthcare providers navigate these new technologies, data on which populations would benefit the most may be critical to inform the most appropriate test for their patient. This study aims to determine whether personal cancer history impacts the rate of detection of PVs that would have been misclassified or undetected with DGT alone.
Methods:Proband clinical history data were curated for all patients undergoing paired DGT/RGT from October 2019 through April 2020. Family history information was not available for this study. We then identified patients meeting National Comprehensive Cancer Network testing criteria based on personal history of cancer alone: breast =Results:Clinical history was provided for 43,145 individuals who underwent concurrent DGT/RGT during the study period. Of these, 4,807 individuals had a PV in one of the 18 genes included in the RNA analysis, 3.2% (n=153) of whom had an RNA-impacted PV. For individuals who met personal history testing criteria (n=10,882), 13.1% had a PV compared to 10.5% of those who did not meet criteria (n=32,263) (OR 1.3; 95% CI 1.2-1.4; pDiscussion:Our findings reveal no difference in the proportion of PVs impacted by RNA evidence in those meeting personal history testing criteria and those who do not. Though further study incorporating family history is needed, these observations suggest that RGT is equally beneficial in patients undergoing DGT regardless of personal history.

PrgmNr 2372 - Role of circulating immune checkpoints in *KIT*-mutated metastatic gastrointestinal stromal tumor (GIST) patients

View session detail

Author Block: **C. Brando**[1], D. Fanale[1], L. Incorvaia[2], L. Algeri[1], A. Bonasera[1], N. Barraco[1], L. Corsini[1], A. Cucinella[1], A. Dimino[1], C. Filorizzo[1], A. Fiorino[1], V. Gennusa[1], G. Madonia[1], L. Magrin[1], E. Pedone[1], M. Ricciardi[1], S. Sammataro[1], R. Sciacchitano[1], R. Scalia[1], G. Badalamenti[1], V. Bazan[2], A. Russo[1]; [1]Dept. of Surgical, Oncological and Oral Sci., Section of Med. Oncology, Univ. of Palermo, Palermo, Italy, [2]Dept. of Biomedicine, NeuroSci. and Advanced Diagnostics (Bi.N.D.), Section of Med. Oncology, Univ. of Palermo, Palermo, Italy

Disclosure Block: **C. Brando:** None.

Circulating immune checkpoints and type of *KIT* mutations have recently been shown to correlate with shorter survival in gastrointestinal stromal tumor (GIST). Our study was aimed to understand if soluble forms of immune checkpoints, such as sPD-1, sPD-L1, sBTN3A1, and pan-sBTN3As, and type of *KIT* mutations may be predictors of survival for metastatic GIST (mGIST) patients, in order to obtain useful information about the clinical evolution of disease.We performed a retrospective study including a cohort of 30 mGIST patients enrolled from February 2015 to June 2017. All patients harboring a KIT exon 11 pathogenic variant (PV) were treated with first-line imatinib 400 mg/day: 16 patients (53.3%) harbored KIT exon 11 deletion or delection/insertion, and 14 (46.7%) carried other PV types (duplication, insertion, or single nucleotide variant). The plasma PD-1, PD-L1, BTN3A1, and panBTN3As concentrations were measured in peripheral blood by specific ELISAs, not yet commercially available.Plasma levels of sPD-1, sPD-L1, sBTN3A1, and pan-sBTN3As, age at diagnosis, and type of *KIT* exon 11 PV were found to be statistically significantly associated with progression-free survival (PFS) in univariable analyses, while in the final multivariable Cox regression model, only the plasma levels of sPD-L1 â¤ 0.7 ng/mL (HR: 0.01; 95% CI: 0.001 to 0.18; p = 0.001) and pan-sBTN3As â¤ 5.0 ng/mL (HR: 4.45; 95% CI: 0.96 to 20.5; p = 0.05), and the absence of *KIT* exon 11 Del or Delins at codons 557 and/or 558 (HR: 0.05; 95% CI: 0.007 to 0.31; p = 0.003) were statistically significant. The absence of *KIT* exon 11 deletions or delins at codons 557 and/or 558 and expression levels of sPD-L1 â¤0.7 ng/mL and pan-sBTN3As â¤ 5.0 ng/mL were independent prognostic factors associated with a longer PFS in mGIST patients harboring a *KIT* exon 11 PV prior to imatinib therapy.Our study revealed that sPD-1, sPD-L1, sBTN3A1, and pan-sBTN3As could be used as prognostic factors in mGIST patients because individuals treated with imatinib with baseline immune checkpoint expression values below the respective thresholds showed improved clinical outcome and longer PFS than those with plasma levels above the cut-offs.

PrgmNr 2373 - Single cell consensus clustering (SC3) with REVEAL™: SingleCell on neuroblastoma and fetal gene expression data

View session detail

Author Block: **S. Sarangi**[1], K. S. Sharma[1], G. Kildisiute[2], Z. Pitluk[1], J. Kinchen[1]; [1]Paradigm4, Waltham, MA, [2]Wellcome Sanger Inst., Cambridge, United Kingdom
Disclosure Block: **S. Sarangi:** Salary/Employment; Paradigm4.
Single-cell RNA sequencing (scRNA-seq) is a crucial technology for dissecting the heterogeneity associated with cancer. Profiling the transcriptome at single-cell resolution enhances identification and characterization of distinct subpopulations of tumor cells for the precision treatment of cancer. Additionally, the relatively recent compilation of reference single-cell atlases such as the Human Cell Atlas of Fetal Gene Expression (HCA-FGE), and the Human Cell Atlas (HCA) provides researchers with an invaluable resource for comparing cells from diseased and normal patients and tissues. Consequently, this can create personalized and more effective interventions by understanding the co-distribution of therapeutic targets. As with any rapidly developing technology, scRNA-seq also faces a number of issues, a couple of which we will address in this study - a) validation of analysis tools such as clustering algorithms, and b) computational platform to perform population-scale analyses. Growth in the number of available scRNA-seq analysis tools has mirrored the growth in sequencing efforts. A critical part of the analysis workflow is clustering of cells into different cell types, and one of the algorithms used is called single-cell consensus clustering (SC3) - noted for its high accuracy. One of the limitations in SC3 is its inability to scale to population-scale datasets due to computational complexity. In this study, we demonstrate implementation of SC3 in our software platform REVEAL™: SingleCell, and apply it to data from a recent publication, which looked at differences and similarities between human neuroblastoma cells and a reference of normal human fetal medullary cells. We also show comparison of the tumor dataset in the aforementioned publication to the HCA-FGE, as well as an HCA dataset on maternal-fetal interface. Our objectives in this study are to address the need for consensus-driven benchmarks for single-cell data and demonstrate the ability of our computational platform to perform population-scale analyses.

PrgmNr 2374 - Targeted NGS-based molecular profiling of conjunctival squamous cell carcinomas

View session detail

Author Block: **F. Y. Demirci**[1], Y-M. Wu[2], A. Bernard[2], L. Axenol[2], A. Chinnaiyan[2], D. R. Robinson[2], H. Demirci[2]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **F.Y. Demirci:** None.
Squamous cell carcinoma (SCC) can rarely occur in conjunctiva, and unlike its common and well-characterized cutaneous counterpart, conjunctival SCC remains understudied and not well understood molecularly. In this study, we conducted targeted next-generation sequencing (NGS)-based genomic and transcriptomic analyses of DNAs and RNAs isolated from fresh-frozen tumor samples obtained from 24 eyes with invasive conjunctival SCC during surgical excision. While our in-depth evaluation of the generated high-throughput data is still underway, our initial findings have shown that the majority of conjunctival SCCs harbored somatic *TP53* mutations, similar to that observed in SCCs affecting other anatomic sites, whereas the *CDKN2A, RB1, NOTCH1/2,* and *PIK3CA* mutations were present less commonly. The most frequently observed outlier gene expression involved *TP63*. While a subset of conjunctival SCCs demonstrated *PD-1/PD-L1* expression, tumor mutation burden was more than 20 Mut/Mb in about half of the cases, suggesting a potential benefit from immunotherapy. Our initial observations suggest that conjunctival SCCs show a number of similar and also some different molecular features as compared to cutaneous SCCs. An improved molecular understanding of conjunctival SCC may facilitate our ability to identify the best targets for precision therapy and improve the clinical outcomes in patients suffering from this invasive malignancy.

PrgmNr 2375 - The landscape and clinical relevance of somatic mutations in a cohort of patients with myelodysplastic syndrome

View session detail

Author Block: **T. Zhang**[1], P. Auer[2], S. Spellman[1], W. Saber[3], Y-T. Bolon[1]; [1]Ctr. for Intl. Blood and Marrow Transplant Res., Natl. Marrow Donor Program/Be The Match, Minneapolis, MN, [2]Natl. Marrow Donor Program/Be The Match, Minneapolis, MN, [3]Ctr. for Intl. Blood and Marrow Transplant Res., Dept. of Med., Med. Coll. of Wisconsin, Milwaukee, WI

Disclosure Block: **T. Zhang:** None.

Myelodysplastic syndromes (MDS) represent a heterogeneous group of myeloid malignancies characterized by diverse genetic alterations in hematopoietic stem cells (HSCs). Although several recent studies have characterized the genomic landscape of patients with MDS, there has yet to be a comprehensive genome-wide survey of somatic mutations and their clinical implications in a sizable cohort. To address this gap, whole blood from 494 patients with MDS was subjected to whole-genome sequencing (at 60X depth) and analysis. We utilized the Octopus pipeline to call somatic variants without matched normal and built a reusable germline variant filtering strategy with a custom user interface. Multiple filters allowed for selection and fine-tuning of criteria, including removal of variants with Gnomad allele frequency above 10x10-06, and removal of noncoding variants in low complexity and repetitive regions, those with no functional indications from ANNOVAR annotations, CADD conservative score under 15, and absence in HGMD or COSMIC database. Across all somatic variants, we observed a mean clone number of 3 (range=1-6) with mean variant allele frequency (VAF)= 25.9%. Recurrent somatic mutations demonstrated lower clonal fractions (mean clone number of 1.5 (range=1-3), mean VAF=19.6%). Among 52 recurrently mutated genes in 252 of 494 MDS cases, *TP53*, *TET2*, *RUNX1*, *DNMT3A*, and *ASXL1* were the most frequently mutated genes in our MDS cohort. We used K-means clustering to stratify our cohort into clinically meaningful sub-groups. Six clusters were identified, including one reference cluster with no recurrent somatic mutations or cytogenetic abnormalities. Two other clusters contained samples with *TP53* mutations and the deletion 5q or monosomy 7 cytogenetic abnormalities. We tested whether the distribution of known risk factors for mortality from MDS differed between these two clusters and the reference sub-group. MDS sub-type was strongly associated with cluster membership (p=3.51e-07), as was presence of complex karyotype (p

PrgmNr 2377 - Using single-cell transcriptomics to fit a multi-level computational model of cancer

View session detail

Author Block: **P. Victori**, F. M. Buffa; Univ. of Oxford, Oxford, United Kingdom
Disclosure Block: **P. Victori:** None.

Computational models can provide mechanistic explanations of biological phenomena and improve data integration and interpretation. In cancer, they can help predict response to targeted and combination therapy. In this project, we focused on triple negative breast cancer (TNBC), a cancer of unmet need.

We have developed an agent-based, multi-level model of tumour spheroid growth where genes, cells and diffusible substances have defined properties and a set of possible actions. Each individual cell in the model contains a gene regulatory network (GRN) that receives inputs from the microenvironment and other cells. This GRN outputs a cell fate such as growth arrest, proliferation or apoptosis. This integration of three different modelling levels - intracellular, intercellular and microenvironment - makes our model capable of robust predictions.

We aimed to use the many publicly available single-cell RNA-seq data to fit our model, which will help predict targets for therapy and elucidate mechanisms of resistance to certain drugs. In particular, we are interested in how TNBC resists EGFR targeted therapy.In order to do that, we imported these datasets into our model by building a probabilistic boolean network where the likelihood of each node being active is relative to the computed enrichment score for each gene/protein, not only to the upstream regulation. After importing the regulatory profile of each cell to fit our model, we observed that the general 'phenotype' of each subpopulation (scRNA-seq cluster) is faithfully reproduced in the model, with the proliferative cluster being the one growing the most, which supports the translatability of the model.

We also imported spatial transcriptomic data produced with several technologies. Our model is capable of importing both the expression data and the spatial data, placing each starting cell in the same position within the model as one cell or spot from the spatial dataset. This way we can simulate clonal competition within the tumour paying attention to its topology and composition.

A next step will be to compare the results of simulating targeted drug treatments with our before and after treatment patient data. We are also working in a way to import pseudotime analyses into our model, by defining a differentiation tree for each subpopulation. Then we will check how mutations and targeted drugs affect these differentiation paths, which is something that has already been observed for some of the patient datasets we are using. We hope that this will provide us with novel mechanistic explanations of drug resistance and tumour growth.

PrgmNr 2379 - WGTS insights in a cancer patient: a tale of two primaries

View session detail

Author Block: **D. Vucenovic**[1], Z. Schlachetzki[2], I. Armogida[1], C. Fielding[1], H. Northen[1], K. D. Robinson[2], F. Kamole[1], M. Bajracharya[1], N. Walker[1], J. Denecker[1], D. McBride[1], T. James[1], Z. Kingsbury[1], M. Mijuskovic[1], J. Ortega[2], D. Eberhard[2], J. Becq[1], M. Ross[1], R. J. Taft[2]; [1]Illumina, Inc., Cambridge, United Kingdom, [2]Illumina, Inc., San Diego, CA
Disclosure Block: **D. Vucenovic:** Salary/Employment; Illumina, Inc.
Sequencing is increasingly used in cancer to guide therapy selection and provide prognostic information. Whole genome sequencing provides a comprehensive view of somatic and germline variants, which may provide clinical benefit and improve patient outcomes. Parallel transcriptome analysis enables investigation of the impact of genomic alternations on gene expression, cancer progression prediction and expression signature-based subtyping.
Here, we describe the genomic and transcriptomic profiling of a patient who was treated for two anatomically distant tumours in an 11-month period. The first was a clear cell renal cell carcinoma (ccRCC). We received resected tumour tissue, as well as adjacent non-malignant tissue, from which we carried out high-depth whole genome (~120x) sequencing on four independent sections and transcriptome sequencing (average of 370 million reads per library) on three sections. We identified somatic small variants, copy number alterations and structural rearrangements characteristic of renal carcinomas, including loss of the chromosome 3p. We also observed intra-tumour heterogeneity, including a *BAP1* mutation with an allele frequency ranging from 0.04 to 0.47. Differential gene expression identified upregulated negative prognostic markers of renal carcinomas.
Shortly afterwards the patient had a bowel resection in management of colorectal cancer, from which we received two spatially separated samples of the tumour and adjacent normal tissue. We again applied high-depth whole genome and transcriptome sequencing and observed somatic variants that have been shown to be frequently mutated in colon adenocarcinomas, including *APC* loss of function mutations. No pathogenic germline variants in mismatch-repair genes were observed, nor were microsatellite instability or signatures of mismatch repair deficiency.
We considered the possibility that the two tumours had a common origin. We found, however, that the genomic rearrangements and characteristic mutations were not shared. Similarly, analysis of the whole transcriptome data in comparison to The Cancer Genome Atlas (TCGA) revealed that the patientâ□□s ccRCC samples clustered with 538 kidney renal clear cell carcinoma (KIRC) samples, and the colon cancer samples clustered with 429 colon adenocarcinoma (COAD) samples. These results confirmed that the patient was suffering from two independent synchronous tumours.
This analysis shows that whole genome and transcriptome sequencing (WGTS) can provide insight into tumour origin and development, and supports its use to guide clinical decision making.

PrgmNr 2380 - Whole-genome sequencing of phenotypically distinct inflammatory breast cancers reveals similar genomic alterations to non-inflammatory breast cancers

View session detail

Author Block: **X. Li**[1], S. Kumar[2], A. Harmanci[3], S. Li[4], R. Kitchen[1], Y. Zhang[5], V. Wali[1], S. Reddy[6], W. Woodward[7], J. Reuben[7], J. Rozowsky[8], C. Hatzis[1], N. T. Ueno[7], S. Krishnamurthy[7], L. Pusztai[1], M. Gerstein[2]; [1]Yale Univ., New Haven, CT, [2]Yale Univ, New Haven, CT, [3]Univ. of Texas Hlth.Sci. Ctr. Houston, Houston, TX, [4]New Haven, CT, [5]Plain City, OH, [6]Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [7]The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [8]YALE Univ., New Haven, CT
Disclosure Block: **X. Li:** None.
Background: Inflammatory breast cancer (IBC) has a highly invasive and metastatic phenotype. However, little is known about its genetic drivers. To address this, we report the largest cohort of whole-genome sequencing (WGS) of IBC cases.
Methods: We performed WGS of 20 IBC samples and paired normal blood DNA to identify genomic alterations. For comparison, we used 23 matched non-IBC samples from the Cancer Genome Atlas Program (TCGA). We also validated our findings using WGS data from the International Cancer Genome Consortium (ICGC) and the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium. We examined a wide selection of genomic features to search for differences between IBC and conventional breast cancer. These include (i) somatic and germline single-nucleotide variants (SNVs), in both coding and non-coding regions; (ii) the mutational signature and the clonal architecture derived from these SNVs; (iii) copy number and structural variants (CNVs and SVs); and (iv) non-human sequence in the tumors (i.e., exogenous sequences of bacterial origin).
Results: Overall, IBC has similar genomic characteristics to non-IBC, including specific alterations, overall mutational load and signature, and tumor heterogeneity. In particular, we observed similar mutation frequencies between IBC and non-IBC, for each gene and most cancer-related pathways. Moreover, we found no exogenous sequences of infectious agents specific to IBC samples. Even though we could not find any strongly statistically distinguishing genomic features between the two groups, we did find some suggestive differences in IBC: (i) The *MAST2* gene was more frequently mutated (20% IBC vs. 0% non-IBC). (ii) The TGF *Î²* pathway was more frequently disrupted by germline SNVs (50% vs. 13%). (iii) Different copy number profiles were observed in several genomic regions harboring cancer genes. (iv) Complex SVs were more frequent. (v) The clonal architecture was simpler, suggesting more homogenous tumor-evolutionary lineages.
Conclusions: Whole-genome sequencing of IBC manifests a similar genomic architecture to non-IBC. We found no unique genomic alterations shared in just IBCs; however, subtle genomic differences were observed including germline alterations in TGFÎ² pathway genes and somatic mutations in the *MAST2* kinase that could represent potential therapeutic targets.

PrgmNr 2381 - A phenome-wide association scan reveals novel associations between genetically increased blood lipid levels and the risk of cholelithiasis

View session detail

Author Block: **S. Kanoni**[1,2], X. Zhu[3,4,5,6], S. E. Graham[7], S. L. Clarke[5,8], K. Bhatti[1], Y. Wang[9], S. Ramdas[10], I. Surakka[7], P. Natarajan[11,12,13,14], Y. V. Sun[15,16,17], C. D. Brown[10], G. M. Peloso[9], C. J. Willer[7,18,19], P. Deloukas[1,2], T. L. Assimes[5,8], Million Veteran Program, Global Lipids Genetics Consortium; [1]Clinical Pharmacology, William Harvey Res. Inst., Barts and the London Med. Sch., Queen Mary Univ. of London, London, United Kingdom, [2]Ctr. for Genomic Hlth., Queen Mary Univ. of London, London, United Kingdom, [3]Dept. of Statistics, The Pennsylvania State Univ., University Park, PA, [4]Huck Inst.s of the Life Sci., The Pennsylvania State Univ., University Park, PA, [5]VA Palo Alto Hlth.Care System, Palo Alto, CA, [6]Dept. of Statistics, Stanford Univ., Stanford, CA, [7]Dept. of Internal Med., Div. of Cardiovascular Med., Univ. of Michigan, Ann Arbor, MI, [8]Dept. of Med., Div. of Cardiovascular Med., Stanford Univ., Stanford, CA, [9]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [10]Dept. of Genetics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [11]Cardiology Div., Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA, [12]Dept. of Med., Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA, [13]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, [14]Cardiovascular Res. Ctr. and Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [15]VA Atlanta Hlth.care System, Decatur, GA, [16]Dept. of Epidemiology, Emory Rollins Sch. of Publ. Hlth., Atlanta, GA, [17]Dept. of BioMed. Informatics, Emory Sch. of Med., Atlanta, GA, [18]Dept. of Computational Med. and Bioinformatics, Univ. of Michigan, Ann Arbor, MI, [19]Dept. of Human Genetics, Univ. of Michigan, Ann Arbor, MI

Disclosure Block: **S. Kanoni:** None.

Blood lipid levels are well documented risk factors for cardiovascular disease but pleiotropic associations might exist for other conditions. Within a trans-ancestry genome-wide association study involving 1.65 million individuals, we identified 2,399 index lipid-associated variants and used the results to develop best performing polygenic risk scores (PRS) for each of the 5 blood lipid traits, including low density lipoprotein (LDL-C), high density lipoprotein (HDL-C), total cholesterol (TC), triglycerides (TG), and non-high density lipoprotein cholesterol (non-HDL-C). We then investigated the association between genetically increased blood lipids and many health-related traits in two large-scale cohorts with deep phenotyping; UK Biobank and the Million Veteran Program. We performed study-wise phenome-wide (PheWAS) scans for each of the lipid PRSs and combined the two datasets via inverse-variance weighted meta-analysis. Our combined dataset for the PRSs scans included results for 739 phecodes and 34 biomarkers. We detected 38 phenotypes associated with the LDL-C PRS at a level of genome-wide significance (GWS), 114 with the HDL-C PRS, 38 with the total cholesterol PRS, 100 with the triglycerides (TG) PRS, and 49 with the non-HDL-C PRS. As expected, multiple coronary and non-coronary atherosclerosis-related phenotypes, as well as hypertension and aortic aneurysm were associated with all five lipid PRSs. All lipid PRSs were also associated at GWS with decreased levels of direct bilirubin; the strongest association was detected for the TC PRS (beta=-0.049 umol/L, SE=0.001, p=1.06e-260). Similarly, all lipid PRSs were associated with low risk for cholelithiasis, apart from the TG PRS, where the direction of effect was reversed. The inverse association between LDL PRS and cholelithiasis had a magnitude of OR=0.937 (95% CIs= 0.923-0.950, p=1.90e-20) in a total of 21,337 cases and 454,050 controls and we further replicated this association in the Michigan Genomics Initiative. In our single-variant by variant PheWAS scan in the UK Biobank, we found 30 unique variants, corresponding to 17 loci, associated with cholelithiasis at GWS. These included variants in four genetic loci (GCKR, ABCG8, CYP7A1 and SULT2A1) previously reported in association with gallstone disease. Our results suggest a substantially larger fraction of lipid-related loci influence the risk of gallstone disease through blood lipid levels than previously

anticipated and may help stratify the risk of subclinical and/or clinical cholelithiasis.

PrgmNr 2382 - Assessing the contribution of rare variants to human complex traits by whole-genome sequencing

View session detail

Author Block: **K. Kundu**[1,2], K. Walter[1], P. Surendran[2], S. Nongmaithem[1], L. Bomba[3], R. Power[3], K. Estrada[4], J. Danesh[2], A. S. Butterworth[2], The INTERVAL Study, N. Soranzo[5]; [1]Wellcome Sanger Inst., Hinxton, United Kingdom, [2]Univ. of Cambridge, Cambridge, United Kingdom, [3]BioMarin, Novato, CA, [4]Biomarin, Novato, CA, [5]Wellcome Trust Sanger Inst, Hinxton/Cambridge, United Kingdom
Disclosure Block: **K. Kundu:** None.
As part of the INTERVAL study (http://www.intervalstudy.org.uk) we have generated whole-genome sequence data (WGS) at 15X coverage for 12,354 samples and whole-exome sequence data (WES) at 50X coverage for 4,070 samples. INTERVAL is a cohort study of approximately 50,000 healthy blood donors, aged 18 years and older, who were consented and recruited from 25 National Health Service Blood and Transplant donor centres across England between 2012 and 2014. WGS and WES data enable the direct calling of rare variants with allele frequencies AFP -8. In the future, we would like to investigate the functional impact on the non-coding, regulatory part of the genome to further enhance our insight into the genetic architecture of complex traits. Mendelian randomization approaches will be employed to establish the causal relationship between the intermediate phenotypes found to be associated in the INTERVAL study and disease outcomes from the PheWAS study.

PrgmNr 2383 - Assessing the role of rare pathogenic variants in heart failure progression in the UK Biobank

View session detail

Author Block: **O. Chazara**[1], Q. Wang[1], S. V. V. Deevi[1], D. S. Paul[1], J-C. Tardif[2,3], M-P. Dube[2,3,4], C. Haefliger[1], K. J. Carss[1]; [1]Ctr. for Genomics Res. (CGR), Discovery Sci., BioPharmaceuticals R&D, AstraZeneca, Melbourn, United Kingdom, [2]Montreal Heart Inst., Montreal, QC, Canada, [3]UniversitÃ© de MontrÃ©al, Faculty of Med., Dept. of medicine, UniversitÃ© de MontrÃ©al, Montreal, QC, Canada, [4]UniversitÃ© de MontrÃ©al Beaulieu-Saucier Pharmacogenomics Ctr., Montreal, QC, Canada
Disclosure Block: **O. Chazara:** Salary/Employment; Astrazeneca.
Heart failure affects ~40 million individuals around the world. In age and sex-adjusted analyses, the relative risk of developing heart failure is 1.69 if one parent is affected and 1.92 if both parents have heart failure, suggesting a genetic component [1].
The majority of heart failure cases are attributed to ischemic heart disease, hypertension, or cardiomyopathy. The latter can be caused by rare, high-impact variants in genes encoding proteins expressed in the heart. Recently, a study highlighted that these rare variants also contribute to the development of heart failure in patients with ischemic heart disease [2]). The role of rare variants in heart failure progression remains to be investigated.
In this study, we tested the hypothesis that rare variants have an impact on heart failure progression using exome sequencing data of ~455,000 participants in the UK Biobank prospective cohort study. We identified up to 11,076 individuals with heart failure and/or cardiomyopathy (based on ICD-10 codes from hospital in-patient diagnoses [3]). We defined â□□progressionâ□□ as cardiovascular death or a composite of cardiovascular death and heart failure hospitalisation. We used Cox Regression with Firth's penalized likelihood to assess the contribution of rare pathogenic variants (collapsed to gene-level) in heart failure progression.
For an endpoint defined as cardiovascular death, highly ranked associations include *KDM3B* (p = 6.89x10-7), *ARHGEF17* (p = 2.0x10-6), and *TTN* (p = 1.04x10-4), but none of these reach the stringent threshold for statistical significance set at p = 2.7x10-7. As these results did not pass the significance threshold, they will need further validation and investigation. We will explore the differences between heart failure patients with and without diagnosed cardiomyopathy, and between patients with and without a diagnosis at the time of enrolment in the UK Biobank.
[1] Lee DS, Pencina MJ, Benjamin EJ, et al. Association of parental heart failure with risk of heart failure in offspring. N Engl J Med. 2006;355(2):138â□□147.
[2] Povysil G, Chazara O, Carss KJ, et al. Assessing the Role of Rare Genetic Variation in Patients With Heart Failure. JAMA Cardiol. 2021:6(4):379-386.
[3] Shah, S., Henry, A., Roselli, C. et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. Nat Commun 2020; 11(1):163.

PrgmNr 2384 - Chromosome X Inactivation is associated with cardiovascular risk and monocyte abundance in an ageing cohort

View session detail

Author Block: **A. L. Roberts**[1], A. Morea[2,1], M. Tomlinson[1], J. S. El-Sayed Moustafa[1], C. Christiansen[1], C. C. Y. Wong[1], T. J. Vyse[1], K. S. Small[1]; [1]King's Coll. London, London, United Kingdom, [2]The FIRC Inst. of Molecular Oncology, Milan, Italy
Disclosure Block: **A.L. Roberts:** None.
Chromosome X inactivation (XCI) is a dosage compensation mechanism which transcriptionally silences one X chromosome to equalise the gene expression of XX females and XY males. The XCI status of a cell is clonally inherited, and though XCI is typically a random process resulting in an equal mosaicism across a tissue, females can display XCI-skew which is the preferential silencing of one X chromosome. XCI-skew has a heritable basis and becomes more prevalent with increasing age in blood tissues; XCI-skew is observed in ~25% of females over 50. XCI-skew is associated with clonal haematopoiesis, which itself has been linked to increased risk of cardiovascular disease, cancer, and all-cause mortality. However, the health risks associated with XCI-skew have not been robustly established. Similarly, mosaic loss of chromosome Y (mLOY) is a heritable cellular phenotype associated with health outcomes in ageing males. Yet it is unknown whether there is a shared genetic susceptibility between these two age-associated traits of sex chromosome dysregulation. We measured XCI cross-sectionally in 1,575 samples from the deeply-phenotyped twin cohort TwinsUK to identify associations with cardiovascular disease risk and genetic susceptibility to mLOY. XCI-skew was assayed from blood-derived DNA (median age = 61) using the Human Androgen Receptor Assay. The Atherosclerosis and Cardiovascular Disease (ASCVD) risk score was calculated, and date-matched full blood count data were available. A polygenic risk score for mLOY was calculated from publicly available summary statistics (Thompson, 2019). Associations were tested using linear mixed effects models, controlling for relatedness and family structure, and relevant covariates including cell-type composition. Similar to clonal haematopoiesis, XCI-skew was positively associated with ASCVD risk score (P=0.01). This was confirmed in XCI-skew discordant twin pairs (N=34), where higher XCI-skew was associated with higher ASCVD risk score (Wilcox paired-test P=0.009) suggesting the association is not driven by shared genetic susceptibility to both traits. XCI-skew was also associated with increased monocyte abundance (P=0.004), an immune cell involved in the inflammatory pathophysiology of cardiovascular disease. However, no association was observed between XCI-skew and the polygenic risk score for mLOY (P=0.89), suggesting limited shared genetic susceptibility between these two traits. Our study is the first to demonstrate an association between cardiovascular disease risk and XCI-skew. Further data from TwinsUK will enable the identification of other age-related health consequences linked to XCI-skew.

PrgmNr 2385 - Discovery of rare variants associated with resting heart rate

View session detail

Author Block: **L. T. Sooriyabandara**[1], J. RamÃrez[2], S. Van Duijvenboden[1], A. Tinker[2], P. B. Munroe[3]; [1]Queen Mary Univ. of London, London, United Kingdom, [2]William Harvey Res. Inst., London, United Kingdom, [3]Queen Mary Univ. of London, London, United Kingdom
Disclosure Block: **L.T. Sooriyabandara:** None.
Resting heart rate (RHR) is an important marker associated with cardiovascular disease, and genome-wide association studies (GWAS) have identified common variants at more than 437 loci. For most loci, the effector genes and relevant biological processes through which HR associations are mediated are not yet fully characterised, identification of rare variants may help to pinpoint candidate genes and explain additional percentage variance of this trait. We performed a rare variant GWAS (RV-GWAS) in 388,223 individuals and whole exome sequencing (WES) analysis in 161,539 individuals for RHR using BOLT-LMM v2.3.5 and determined the functional impact of rare variants using Variant Effect Predictor. All participants were of European ancestry from UK Biobank. We used PLINK 2.0 to perform genotype quality control, including a minor allele frequency (MAF) > 6.2×10^{-6} and -12. We first estimated the heritability of RHR in the RV-GWAS dataset, rare variants contributed 1.95% of trait variance. This is approximately $1/10^{th}$ of the heritability estimated from common variants. The RV-GWAS identified rare variants at 29 loci, 20 at novel loci and nine at previously reported RHR loci. The effect sizes ranged between 0.81 and 5.75 beats per minute per allele (a variant at the *NEO1* locus had the largest effect size, variants in this gene have suggestive association with heart rate variability). At novel loci, we observe three missense variants in the genes *AKTIP*, *TBX5* and *DBH*. *AKTIP* encodes an *AKT* interacting protein, and a knockout mouse model has heart abnormalities. The rare variants at *TBX5* and *DBH* are identical to rare variants recently reported for blood pressure traits. Rare variants at previously reported RHR loci include several well-established cardiovascular genes (*MYH6*, *CACNA1D*, *GJA1*). Mutations in these genes may cause deficits in autonomic and cardiovascular function, leading to sinoatrial node dysfunction and cardiomyopathy. The WES analysis identified rare variants at seven loci. Five loci do not overlap results from the RV-GWAS, 3 are novel and 2 are at previously reported loci. At novel loci, there are two missense variants in genes *SPATA31A1,* and *HIGD1B*. There are no prior associations described with cardiovascular phenotypes for both genes. At the third novel locus, the nearest gene is *TBC1D32*. Several cardiovascular abnormalities have been observed in a knockout mouse model supporting its candidature. In summary, we found 34 rare variants associated with RHR, our findings provide new biological mechanisms mapping to cardiovascular risk and novel candidates for functional experimentation.

PrgmNr 2386 - Epigenetics of Single and Multisite Atherosclerosis in African Americans from the Genetic Epidemiology Network of Arteriopathy (GENOA)

View session detail

Author Block: **F. Ammous**[1], W. Zhao[1], L. Lin[1], S. Ratliff[1], T. Mosely[2], L. Bielak[1], X. Zhou[3], P. peyser[4], S. L. Kardia[4], J. A. Smith[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Memory Impairment and Neurodegenerative Dementia (MIND) Ctr., Univ. of Mississippi Med. Ctr., Jackson, MS, [3]Univ MICHIGAN, Ann Arbor, MI, [4]Univ Michigan, Ann Arbor, MI
Disclosure Block: **F. Ammous:** None.
Background: DNA methylation, an epigenetic mechanism modulated by lifestyle and environmental factors, may be an important biomarker of complex diseases including cardiovascular diseases (CVD) and subclinical atherosclerosis. **Methods:** DNA methylation in blood samples from 391 African Americans from GENOA was assessed at baseline and atherosclerosis was assessed five and 12 years later. Using linear mixed models, we examined the association between previously-identified CpGs for coronary artery calcification (CAC) and carotid plaque in another study, both individually and aggregated into methylation risk scores (MRS_{CAC} and $MRS_{cartoid}$), and four measures of atherosclerosis (CAC, abdominal aorta calcification (AAC), ankle brachial index (ABI), and multisite atherosclerosis based on gender-specific quartiles of the single site measures. We also examined the association between four epigenetic age acceleration measures (IEAA, EEAA, PhenoAge acceleration, and GrimAge acceleration) and the four atherosclerosis measures. Finally, we characterized the temporal stability of the epigenetic measures using repeated DNA methylation at five years from baseline (N=193). **Results::** After adjusting for traditional CVD risk factors, one and six CpGs were associated with AAC and multisite atherosclerosis at false discovery rate $_{carotid}$ was associated with 1.6-fold increase in the Agatston score of CAC (95%CI 1.09-2.40) and AAC (95%CI 1.13-2.68), and a 0.7 units (95%CI 0.21-1.13) increase in multisite atherosclerosis, after adjusting for CVD risk factors. MRScarotid explained 5.3%, 2.7%, and 5.5% of the variability of CAC, AAC, and multisite atherosclerosis. A 5-year increase in GrimAge acceleration (~1 SD) was associated with a 1.6-fold (95%CI 1.11-2.25) increase in the Agatston score of AAC and 0.7 units (95%CI 0.33-1.07) increase in multisite atherosclerosis, all after adjusting for CVD risk factors. All epigenetic measures were relatively stable over five years, with the highest intraclass correlation coefficients observed for $MRS_{carotid}$ and GrimAge acceleration (0.82 and 0.89, respectively). **Conclusions:** We found evidence of an association between DNA methylation and atherosclerosis at multiple vascular sites in a sample of African Americans. These findings deepen our understanding of the relationship between biological aging and atherosclerosis and suggest that further evaluation of these potential biomarkers is warranted.

PrgmNr 2387 - Estimating the causal effects of cardiometabolic factors on coronary artery disease in British Pakistanis and Bangladeshis: A trans-ancestry Mendelian Randomisation study

View session detail

Author Block: **D. Dunca**[1], Q. Huang[2], N. Sallah[1], H. Martin[2], T. Lumbers[1], K. Kuchenbaecker[1]; [1]Univ. Coll. London, London, United Kingdom, [2]Wellcome Sanger Inst., Cambridge, United Kingdom
Disclosure Block: **D. Dunca:** None.
British people with South Asian (SAS) ancestry have a higher risk of coronary artery disease (CAD) than other ancestry groups. However, genetic research into the causes of CAD has focused on primarily European (EUR) ancestry individuals. Statistical power can be the limiting factor when extending Mendelian Randomisation (MR) analyses to non-European populations because independent ancestry matched GWAS for risk factors of interest might not be sufficiently large. Here we compared different strategies for trans-ancestry MR and used them to assess the causal effect of cardiometabolic risk factors (BMI, triglycerides, HDL-cholesterol, LDL-cholesterol, systolic and diastolic blood pressure) on the risk of CAD in 22,000 British Pakistani and Bangladeshi (BPB) individuals from the Genes&Health (G&H) cohort. First, we followed the standard MR approach of using an independent ancestry matched sample to derive instruments: a two-sample MR of CAD in G&H with summary statistics for risk factors from the SAS group in UK Biobank. However, insufficient numbers of genome-wide significant instruments were identified for the exposure variables in the UK Biobank SAS population for a well powered MR analysis, due to limited sample size (~8,000). We assessed different strategies to address this. First, we used a less stringent p-value threshold (p

PrgmNr 2388 - Exploring causal relationships between cardiometabolic traits and cancers using pathway-specific genetic instruments

View session detail

Author Block: **Z. Balkhiiarova**[1,2], A. Ulrich[1], A. Demirkan[1], L. Zudina[1], I. Pupko[1], J. G. Maina[3], P. Froguel[3,2], M. Gunter[4], M. Kaakinen[1,2], I. Prokopenko[1,3,5]; [1]Dept. of Clinical & Experimental Med., Univ. of Surrey, Guilford, United Kingdom, [2]Dept. of Metabolism, Digestion and Reproduction, Imperial Coll. London, London, United Kingdom, [3]UMR 8199 - EGID, Inst. Pasteur de Lille, CNRS, Univ. of Lille, Lille, France, [4]Nutrition and Metabolism Section, Intl. Agency for Res. on Cancer, World Hlth.Organization, Lyon, France, [5]Inst. of Biochemistry and Genetics, Ufa Federal Res. Ctr. Russian Academy of Sci., Ufa, Russian Federation

Disclosure Block: **Z. Balkhiiarova:** None.

Introduction: Epidemiological studies suggest that people with type 2 diabetes (T2D) are at increased risk for various cancers. T2D and postmenopausal breast (BrC), colorectal, prostate and pancreatic cancers share many risk factors, but potential biological links between them are incompletely understood. We aimed to investigate causality between biological pathways underlying T2D, glycaemic traits and cancers using clustering and Mendelian Randomization (MR). Materials and Methods: We first used agglomerative hierarchical clustering to group 1,083 genetic variants associated with T2D, glycaemic traits and four cancers based on their effects on 35 metabolic/inflammatory/tumour- and obesity-related phenotypes. We obtained 5 groups in total. The adiposity cluster comprised 464 variants, including 36 with pronounced effects on triglycerides and age at menarche, respectively. The beta-cell function cluster grouped 96 variants with shared effects on hormone and lipid levels. The sex hormones cluster included 409 variants with lower levels of sex hormone binding globulin and testosterone. The metabolic syndrome cluster comprised 78 T2D SNPs with higher insulin resistance.We then applied a two-sample MR framework to investigate the role of these four pathways in cancer development. Effect estimates for the same or proxy variants (r^2>0.8) on T2D and cancers were obtained from largest-to-date GWAS.Results: The adiposity pathway showed evidence of a negative causal relationship between BMI and BrC (OR_{MR}=0.66, P-value=4.22x10^{-5}) mediated by later age at menarche. Further analysis using 409 BMI variants as instruments also suggested a causal relationship between adiposity and BrC ($\hat{I}^2_{MR}$=0.88, P-value= 5.59x10^{-5}).Conclusions: Dissection of T2D risk variants into distinct pathways improved our ability to detect causal relationships between specific T2D pathways and cancers. Adiposity and age at menarche were intermediate traits underlying the observed relationship between BrC and T2D.

PrgmNr 2389 - Gene-SCOUT: gene-based biomarker signatures can assist identification of novel genes from phenome-wide association analyses

View session detail

Author Block: **L. Middleton**[1], Q. Wang[1], A. R. Harper[1], A. Nag[1], A. Reznichenko[2], D. Vitsios[1], S. Petrovski[3,1]; [1]Ctr. for Genomics Res., Discovery Sci., BioPharmaceuticals R&D, AstraZeneca, Cambridge, United Kingdom, [2]Translational Sci. & Experimental Med., Res. and Early Dev., Cardiovascular, Renal and Metabolism, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden, [3]Univ. of Melbourne, Melbourne, Australia
Disclosure Block: **L. Middleton:** Salary/Employment; AstraZeneca.
Large-scale PheWAS, performed using densely-phenotyped cohorts such as the UK Biobank (UKB), reveal many statistically robust gene-phenotype relationships. We hypothesised that we can leverage these association statistics to construct â□□quantitative trait fingerprintsâ□□ per human gene to reflect an individual geneâ□□s biomarker statistics and then identify gene groups in the human exome on the basis of similar fingerprints to achieve a holistic view of their impact on human biomarkers. Here, we present Gene-SCOUT (Gene-Similarity from COntinUous Traits), a tool that identifies genes with similar quantitative trait fingerprints to a gene of interest. A fingerprint reflects the collection of biomarkers identified to be statistically associated with a gene of interest (hyperparameter recommended setting at p-5) across multiple qualifying variant models and with similar effect directions (cosine similarity measure). These similarity metrics allow highly similar genes to be clustered together. The underlying gene-biomarker population-scale association statistics were obtained from a gene-level rare variant collapsing analysis performed on ~1500 quantitative traits using ~455K UK Biobank exomes, including the recently released Nightingale metabolomic features, available for ~120K of the UKB individuals.
This novel similarity score both helps to expand our understanding of the key biological systems a gene is involved in, and can also suggest alternative drug targets for known non-tractable genes. Using *APOB* as an exemplar, within the top ten ranked genes identified to have the most comparable fingerprint to *APOB*, five are **known cholesterol-lowering** targets **(*PCSK9, ANGPTL3, APOC3, PDE3B and ABCA1)***, thereby demonstrating the robustness of the approach while also suggesting alternative genes that when genetically aberrated in humans have similar biomarker fingerprints. To support the resulting similarities, we provide enrichment analyses based on neighbouring genes to establish whether gene clusters are significantly enriched for Gene Ontology biological processes or human disease traits.
Gene-SCOUT can facilitate the discovery of novel genes beyond those reported in conventional PheWAS analyses and enables them to be assessed against established disease targets. We also developed a web resource to allow users to either prespecify a query gene and visually explore a network of similar genes, according to their quantitative trait fingerprints, or define a list of specific traits and directionalities to retrieve genes satisfying the desired multi-biomarker fingerprint.

PrgmNr 2391 - Polygenic risk and its interaction with lifestyle for cardiovascular mortality

View session detail

Author Block: **J-S. Yun**[1], S-H. JUNG[2], M. Shivakumar[3], B. Xiao[4], W-Y. Park[5], A. V. Khera[6], H-H. Won[7], D. Kim[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]The Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [3]Philadelphia, PA, [4]Palo Alto, CA, [5]Samsung Med Ctr, Seoul, Korea, Republic of, [6]Massachusetts Gen. Hosp., Boston, MA, [7]Sungkyunkwan Univ, Seoul, Korea, Republic of
Disclosure Block: **J. Yun:** None.
Background Previous studies primarily targeted the ability of polygenic risk scores (PRSs) to predict a specific disease, and only a few studies have investigated the association between genetic risk scores and cardiovascular (CV) mortality. We assessed PRSs for coronary artery disease (CAD) and type 2 diabetes (T2D) as predictive factors for CV mortality. **Methods** We used genetic and phenotypic data from UK Biobank participants aged 40-69 years at baseline, collected with standardized procedures. Genome-wide PRSs were constructed using >6 million genetic variants. Cox proportional hazard models were used to analyze the relationship between PRS and CV mortality with stratification by age, sex, disease status, and lifestyle behavior. **Results** Of 377,909 UK Biobank participants, 3,210 (0.8%) died due to CV disease during a median follow-up of 8.9 years. CV mortality risk was significantly associated with CAD PRS (low vs. very high genetic risk groups, CAD PRS hazard ratio [HR] 2.61 [2.02-3.36]) and T2D PRS (HR 2.08 [1.58-2.73]), respectively. These relationships remained significant even after adjustment for demographic and clinical factors. In the very high genetic risk group, adherence to an unfavorable lifestyle was further associated with a substantially increased risk of CV mortality (favorable versus unfavorable lifestyle with very high genetic risk for CAD PRS, HR 8.31 [5.12-13.49]; T2D PRS, HR 5.84 [3.39-10.04]). There was no evidence of significant interaction between PRSs and age, sex, or lifestyle behavior in predicting the risk of CV mortality. **Conclusions** PRSs for CAD or T2D and lifestyle behaviors are independent predictive factors for future CV mortality in the white, middle-aged population. PRS-based risk assessment could be useful to identify individuals who need intensive behavioral or therapeutic interventions to reduce the risk of CV mortality.

PrgmNr 2392 - Polygenic score from a large GWAS predicts cases of heart failure with reduced ejection fraction (HFrEF) but not preserved ejection fraction (HFpEF)

View session detail

Author Block: **K-H. Wu**[1], N. J. Douville[2], B. N. Wolford[1], I. Surakka[1], W. E. Hornsby[1], M. C. Konerman[2], M. R. Mathis[2], M. J. Daly[3], B. M. Neale[3], C. J. Willer[1], W. Zhou[4], the Global Biobank Meta-analysis Initiative; [1]Univ. of Michigan, Ann Arbor, MI, [2]Michigan Med., Ann Arbor, MI, [3]Massachusetts Gen. Hosp., Boston, MA, [4]Massachusetts general Hosp., Broad Inst., Boston, MA
Disclosure Block: **K. Wu:** None.
Introduction Heart failure (HF) is a complex disease with multiple subtypes that have distinct etiologies, pathophysiologies, and genetic risk. The performance of a heart failure polygenic score (PGS), derived from a new genome-wide association study (GWAS), was tested to predict cases of heart failure with reduced ejection fraction (HFrEF) and preserved ejection fraction (HFpEF). **Methods** GWAS was performed within the Global Biobank Meta-analysis Initiative (GBMI) â a global collaboration among 13 biobanks across the world with diverse ancestries. GWAS summary statistics were then used to generate a PGS in a combined cohort from the Michigan Genomics Initiative and Cardiovascular Health Improvement Project (MGI/CHIP). Heart failure cases in the GBMI training dataset were defined based upon ICD codes, which did not distinguish between HF subtypes. Electronic health record data available within MGI/CHIP enabled further classification of patients into HFrEF and HFpEF, using a previously validated methodology incorporating ICD diagnoses, free-text language processing, and left ventricular ejection fraction (LVEF). Twenty percent of the cases derived with this method were further adjudicated by clinicians to confirm the phenotype. To compare the predictive ability of PGS, we evaluated logistic regression models with PGS adjusted for age, sex, and principal components derived from genotype data separately for both HFrEF and HFpEF phenotypes. **Results** In the GBMI training dataset, genetic data was analyzed from a total of 67,049 HF patients from 1,305,592 samples from 6 ancestral populations: 25.4% of the samples were of non-European ancestry. The GWAS identified 22 index variants that reached genome-wide significance. The MGI/CHIP validation dataset contained: 360 HFrEF patients, 232 HFpEF patients, and 24,313 healthy controls. The genome-wide PGS is a significantly better predictor of HFrEF compared to HFpEF in the MGI/CHIP cohort. In the HFrEF model, the PGS had an adjusted odds ratio (aOR) of 1.40 (95% CI: 1.25-1.57; p-value: 1.25×10^{-8}) compared to an aOR of 1.08 (95% CI: 0.93-1.24; p-value: 0.30) in the HFpEF model. **Conclusion** Our analyses showed that a PGS for heart failure derived from GBMI data is useful in predicting HFrEF in an independent dataset. The difficulty in predicting HFpEF could result from: (i) the GBMI HF phenotype preferencing HFrEF over HFpEF, (ii) increased diagnostic accuracy in HFrEF in the evaluation cohort (due to inclusion of decreased LVEF), or (iii) greater genetic heterogeneity in the HFpEF population. Future studies focused on a GWAS for HFpEF may create a more useful polygenic score, if the trait is sufficiently heritable.

PrgmNr 2393 - The impact of polyunsaturated fatty acids biosynthesis on the risk of cardiovascular diseases: A Mendelian randomization study in up to 1,153,768 European ancestry individuals

View session detail

Author Block: **M. Borges**[1], P. Haycock[1], J. Zheng[1], G. Hemani[1], L. Howe[1], A. D. Hingorani[2], D. A. Lawlor[1]; [1]Univ. of Bristol, Bristol, United Kingdom, [2]Univ. Coll. London, London, United Kingdom
Disclosure Block: **M. Borges:** None.
Background: Despite early interest in the cardiovascular effects of polyunsaturated fatty acids (PUFA), the evidence linking PUFA to cardiovascular diseases (CVDs) is still controversial. Genetic variants regulating fatty acid desaturases genes (e.g. *FADS1*), coding for rate-limiting enzymes in PUFA biosynthesis (e.g. D5D), can be used as causal anchors to investigate the involvement of PUFA in CVD aetiology.
Aims: We used Mendelian randomization (MR) to explore the effect of higher D5D activity on a wide range of CVDs in up to 1,153,768 European ancestry individuals. In addition, we explored the three key scenarios that could lead to spurious MR findings (i.e. horizontal pleiotropy, population structure, and selection bias).
Methods: We used summary-data MR to investigate the effect of higher D5D activity (proxied by the ratio of arachidonic acid to dihomo-gamma-linolenic acid) on CVDs risk using rs174546 as the genetic instrument. To assess the plausibility of bias by horizontal pleiotropy, we used genetic colocalization and multivariable Mendelian randomization jointly modelling the expression of multiple genes within the *FADS1* locus. To assess confounding by residual population structure, we compared the association of rs174546 with established CVDs risk factors (i.e. LDL-cholesterol, triglycerides, systolic blood pressure, glycated haemoglobin, smoking, and body mass index) between unrelated individuals and within-siblings (up to 68,691 sibships). To explore selection bias, we carried out a positive control Mendelian randomization analyses of established risk factors on CVDs risk.
Results: Our main findings suggest that higher D5D activity is related to higher risk of coronary artery disease, ischemic stroke, heart failure, atrial fibrillation, peripheral artery disease, venous thromboembolism, and aortic valve stenosis. Multivariable MR confirmed that main findings were driven by higher *FADS1* expression (rather than by the expression of other genes in the region) in multiple tissues and genetic colocalization pointed out a shared genetic variant between genetic signals for D5D activity and LDL-cholesterol. The association of rs174546 with CVDs risk factors was broadly similar when comparing unrelated individuals and siblings. In the positive control analyses, we observed the expected effect of risk factors on CVDs risk.
Conclusions: MR findings indicate that lifelong exposure to higher D5D activity is related to higher risk of several CVDs among Europeans. Sensitivity analyses support this interpretation and indicate LDL-cholesterol as a potential mediating trait between PUFA biosynthesis and CVDs risk.

PrgmNr 2394 - The transcriptional landscape of human atria and its role in atrial fibrillation: the RACE-V Consortium

View session detail

Author Block: **A. Isaacs**[1], B. Mulder[2], M. Gilbers[1], M. Kawczynski[1], E. Bidar[1], B. Maesen[1], J. Maessen[1], J. Winters[1], E. Dâ□□Alessandro[1], B. Scaf[1], M. Rienstra[2], I. C. Van Gelder[2], H. J. G. Crijns[1], M. Stoll[1], U. Schotten[1], RACE-V Consortium; [1]Maastricht Univ., Maastricht, Limburg, Netherlands, [2]Univ. of Groningen, Groningen, Netherlands
Disclosure Block: **A. Isaacs:** None.
Atrial fibrillation (AF), a complex, multifactorial disease, is the most common form of arrhythmia and predisposes patients to several adverse outcomes. Heterogeneity in individual characteristics (e.g., surgical indication and comorbidities) can confound associations between AF and atrial expression. To counter this, carefully matched groups (n = 15 per group; SR: no history of AF; parAF: paroxysmal AF; and persAF: persistent AF) were selected from the RACE-V study and utilized to study atrial expression levels. Unmatched samples were used for replication.
Left (LA) and right atrial (RA) appendages were biopsied and detailed phenotypic information was collected. RNA was extracted and poly-A tailed molecules were sequenced (paired-end, 75 base pair, with Illumina chemistry on a NextSeq500). Raw data was aligned to the human genome (GRCh38p13) and quantified using STAR. Differential expression (DE) analyses were performed with DESeq2. All other statistics were implemented in R.
In total, 29 DE transcripts were observed in LA and 16 in RA for the comparison of persAF and SR. Seven of these were present in both atria. In the replication sample (n = 68 LA and n = 63 RA), most effects were concordant in direction (LA: 29/29, 100%; RA: 15/16, 94%) and were nominally significant (LA: 24/29, 83%; RA: 15/16, 94%). Majorities were replicated at transcriptome-wide significance (LA: 19/29, 66%; RA: 13/16, 81%). No DE was observed for parAF vs. SR or persAF vs. parAF.
DE persAF transcripts were also investigated in parAF. They were greatly enriched for both concordant effect directions (LA: 28/29, 97%, $P = 1.1 \times 10^{-7}$ and RA: 16/16, 100%, $P = 3.5 \times 10^{-5}$) and nominal significance levels (LA: 22/29, 76%, $P = 2.6 \times 10^{-23}$ and RA: 10/16, 63%, $P = 5.9 \times 10^{-10}$). Typically, effects for parAF were in between those for persAF and SR.
In the persAF vs. SR comparison, 100% of significant LA and RA transcripts had a concordant effect direction in the other tissue. LA DE transcripts were highly enriched for nominal significance in RA (25/29, 86%, $P = 5.8 \times 10^{-29}$), and vice versa (14/16, 88%, $P = 6.7 \times 10^{-17}$), with highly correlated effect estimates.
Although poly-A tailed RNA does not capture all non-coding RNA, many such transcripts were identified, suggesting that transcriptional regulation plays an important role in AF. Two of these were anti-sense transcripts for crucial calcium handling molecules *CALM1* and *CALM3*. Also of note is the finding that persAF DE transcripts showed substantial evidence of association with parAF. These effects were typically between persAF and SR, in line with the broadly accepted notion of AF as a progressive disease. Other transcripts identified include *RALGPS1*, *MDM1*, and *PPIB*.

PrgmNr 2395 - A glomerular transcriptomic landscape of *APOL1* in Black patients with focal segmental glomerulosclerosis

View session detail

Author Block: **M. T. McNulty**[1], D. Fermin[2], F. Eichinger[2], D. Jang[3], N. P. Burtt[3], M. Kretzler[2], M. Pollak[4,5], J. Flannick[1,3], D. Friedman[4,5], M. Sampson[1,5,3]; [1]Boston Children's Hosp., Boston, MA, [2]Univ. of Michigan - Michigan Med., Ann Arbor, MI, [3]Broad Inst., Cambridge, MA, [4]Beth Israel Deaconess, Boston, MA, [5]Harvard Med. Sch., Boston, MA

Disclosure Block: **M.T. McNulty:** None.

Apolipoprotein L1 (*APOL1*)-associated focal segmental glomerulosclerosis (FSGS) is the dominant form of FSGS in Black people. There are no targeted therapies for this condition, in part because the molecular mechanisms underlying *APOL1*â□□s pathogenic contribution to FSGS are incompletely understood. Studying the transcriptomic landscape of *APOL1* FSGS in patient kidneys is an important way to discover genes and molecular behaviors that are unique or most relevant to the human disease. With the hypothesis that the pathology driven by the high-risk (HR) *APOL1* genotype is reflected in alteration of gene expression across the glomerular transcriptome, we compared expression and co-expression profiles of 15,703 genes in 16 Black FSGS patients with a HR vs 14 with a low-risk (â□□LRâ□□) *APOL1* genotype. Expression data from *APOL1*-inducible HEK293 cells and normal human glomeruli were used to pursue genes and molecular pathways illuminated in these studies. We discovered (1) increased expression of *APOL1* in HR and nine other significant differentially expressed genes, including stanniocalcin (*STC1*), which has a role in mitochondrial and calcium-related processes, (2) differential correlations between HR and LR *APOL1* and metabolism pathway genes, but similar correlations with extracellular matrix- and immune-related genes, (3) significant loss of co-expression of mitochondrial genes in HR FSGS, and (4) an NF-kB -down-regulating gene, *NKIRAS1*, as the most significant hub gene with strong differential correlations with NDUF family and immune-related genes. Overall, differences in mitochondrial gene regulation appear to underlie many differences observed between HR and LR FSGS. All data are available for secondary analysis through the â□□*APOL1* Portalâ□□ (http://APOL1portal.org).

PrgmNr 2396 - Assessment of the Genetic Risk Scoring Algorithm GRS2 for Type 1 Diabetes in African American Children

View session detail

Author Block: **J. Qu**[1], H-Q. Qu[2], Y. Liu[3], F. D. Mentch[4], J. Glessner[5], X. Chang[6], M. March[7], J. Li[2], J. J. Connolly[2], J. D. Roizen[8], P. M. Sleiman[6], H. Hakonarson[7]; [1]Childrenâ□□s Hosp. of Philadelphia, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]The children's Hosp. of philadelphia CHOP, Philadelphia, PA, [4]Childrens Hosp. of Philadelphia, Philadelphia, PA, [5]Children s Hosp. of Philadelphia, Philadelphia, PA, [6]CHOP, Philadelphia, PA, [7]Children S Hosp. of Philadelphia, Philadelphia, PA, [8]The Children S Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **J. Qu:** None.

Type 1 diabetes (T1D) is caused by autoimmune destruction of pancreatic Î²-cells, which is most prevalent in children with European ancestry, but also presents a serious burden among African American (AA) children (Mayer-Davis, Beyer et al. 2009). Once diagnosed, the disease progress is irreversible, and the patients will depend on insulin therapy. Precise risk prediction to enable prevention or early intervention is therefore warranted. The genetic risk scoring (GRS) system for T1D, i.e. T1D-GRS2 (Sharp, Rich et al. 2019), using 67 SNPs from known autoimmune loci, demonstrated excellent performance for the prediction of T1D in European children. With genetic susceptibility of T1D varying significantly across human populations, our study aimed to assess the performance of T1D-GRS2 in AA children. We investigated 168 T1D AA cases versus 1366 non-diabetes AA controls. In comparison, 361 T1D cases of European American (EA) versus 1943 non-diabetes EA controls were also studied. The patients were recruited by the Center for Applied Genomics (CAG) at CHOP, which has established a large pediatric biobank coupled to comprehensive electronic medical record (EMR). The genotyping was done with the Illumina Genotyping BeadChips with at least 550,000 SNPs genotyped. Genome-wide imputation was done with the TOPMed Imputation Server using the TOPMed (Version R2 on GRC38) Reference Panel. The *HLA* region was additionally imputed, using the SNP2HLA software (Jia, Han et al. 2013). The population ancestry of each individual was validated by principal component analysis (PCA) with genome-wide SNP markers. As shown by our results, the T1D AA cases had lower GRS2 scores than the EA cases (8.13Â±2.33 *vs.* 10.52Â±2.20, *P*=3.62E-27), while the AA controls had also lower GRS2 scores than the EA controls (5.24Â±2.32 *vs.* 7.41Â±2.53, *P*= 8.86E-127). The GRS2 had AUC (95% CI)=0.807 (0.779, 0.835) to predict T1D in the CAG AA cohort, compared to AUC (95% CI)=0.823 (0.804, 0.842) in the CAG EA cohort. The prediction of T1D has a sensitivity of 0.613 and a specificity of 0.834 with the maximum Matthews correlation coefficient (MCC) at the cutoff of GRS2=7.43 in AA, compared to a sensitivity of 0.623 and a specificity of 0.833 at the cutoff of GRS2= 9.75 in EA. Furthermore, we attempted to improve the GRS2 performance in AA by including 4 more SNPs (GRS2â□□, *HLA-DQ*: rs9273363; non-*HLA*: rs926169, rs10788599, and rs56380902) identified with T1D association in AA. The GRS2â□□ has an improved AUC (95% CI)= 0.826 (0.800, 0.852) in the CAG AA cohort. Our study suggests that the T1D GRS2 is applicable to the AA population, and could be further improved by including more SNP markers associated with AA T1D.

PrgmNr 2397 - Causal association between circulating metabolites and age at menarche: A two-sample Mendelian randomization study

View session detail

Author Block: **M. Yazdanpanah**[1], N. Yazdanpanah[1], D. Manousaki[1,2]; [1]Res. Ctr. of the Sainte-Justine Univ. Hosp., Univ. of Montreal, Montreal, QC, Canada, [2]Dept.s of Pediatrics, Biochemistry and Molecular Med., Univ. of Montreal, Montreal, QC, Canada
Disclosure Block: **M. Yazdanpanah:** None.
Pubertal timing, as reflected in the age at menarche (AAM) in girls, has a multifactorial etiology and a polygenic basis, but the biological mechanisms underlying variation in the AAM are not completely understood. It has been suggested that earlier pubertal maturation may be associated with adverse health outcomes. Thus, recognizing early predictors of precocious puberty could enhance our understanding of its pathophysiology and identify potential biomarkers for screening and diagnosis. Here, we applied a two-sample Mendelian Randomization (MR) study to screen thousands of previously measured circulating metabolites for a causal association with AAM. We identified 903 unique circulating metabolites with genome-wide significant single nucleotide polymorphisms (SNPs) in four large scale metabolomic genome-wide association studies (GWAS), as instruments in our MR study. Among these 903 molecules, genetically altered levels of 17 metabolites were associated with AAM in ~370,000 women from the ReproGen Consortium GWAS, after correction for multiple testing: specifically, 4 amino acids (isolucine threonine, glutamate, and histidine); 8 lipids (phosphatidylcholines and lyso-phosphatidylcholines), 3 carbohydrates (mannose, Inositol, myo-inositol) and 2 metabolites of unknown class (all Wald ratio p-values -3). Two metabolites with multiple SNPs showed evidence for causal relationships with AAM in the main analysis and in pleiotropy-robust sensitivity MR analyses, namely lyso-phosphatidylcholine and imidazole lactate (p-value inverse variance-weighted (IVW) MR $=1.23x10^{-3}$, and $5.10x10^{-2}$, respectively). Further, we tested whether the prioritized metabolites from our MR study for AAM was also associated with the Tanner stage for breast or testicular development or time of pubertal growth spurt in children from the Early Growth Genetics Consortium GWAS (N= 6,147 females and N= 3,769 males). 2-methylmalonyl carnitine was significantly associated with Tanner stage in males (Wald ratio p-value= $8.16 x10^{-7}$), while glutamate showed an association with both Tanner stage and pubertal growth (Wald ratio p-value for both = $4.0 x10^{-3}$). Our MR study provides evidence supporting a causal role of genetically altered levels of specific blood metabolites in pubertal timing. Our results are in line with previous studies that have linked blood glutamate levels with plasma estrogen and progesterone levels. Our findings support the presence of differences in the metabolic profile of children with altered pubertal timing, while they pinpoint promising biomarkers and/or potential therapeutic targets for early puberty.

PrgmNr 2398 - Discovery of Type 2 Diabetes genes using an accessible tissue

View session detail

Author Block: **D. Davtian**[1], J. M. Schwenk[2], M. McCarthy[3], A. Mahajan[4], M-G. Hong[5], E. T. Dermitzakis[6], H. Im[7], A. ViÃ±uela[8], E. Pearson[1], A. Brown[1]; [1]Univ. of Dundee, Dundee, United Kingdom, [2]KTH Royal Inst. of Technology, Stockholm, Sweden, [3]Univ. of Oxford, Oxford, United Kingdom, [4]Genentech, San Francisco, CA, [5]Karolinska Inst, Stockholm, Sweden, [6]Univ. of Geneva, Geneva, Switzerland, [7]The Univ. of Chicago, Chicago, IL, [8]Univ. of Geneva Med. Sch., Geneva, Geneva, Switzerland

Disclosure Block: **D. Davtian:** None.

Introduction

Transcriptome Wide Assocation Study (TWAS) methods use reference expression datasets to link genes to disease. Ideally, the reference dataset should use samples from a tissue relevant for the disease, but many relevant tissues are difficult to collect, limiting sample size. However, from projects such as GTEx we have observed that many genetic signals are shared across multiple tissues. Here we investigate the power to discover Type 2 Diabetes (T2D) relevant genes using a well powered whole blood expression reference.

Materials and Methods

Genotypes, gene expression and protein levels data from DIRECT consortium (n = 3029 pre-diabetics and diabetics, 16,205 genes and 452 proteins) from whole blood samples were used to train predictive models using the PrediXcan, Elastic Net approach. These models were then combined to GWAS summary statistics from the DIAGRAM consortium (n = 898,130) and used to calculate gene-trait association scores with S-PrediXcan. We compared our results to the publicly available GTEx v8 PrediXcan results (49 tissues, maximum sample size 706).

Results

We found 160 significant associations between T2D and predicted gene expression levels and 4 with predicted protein levels. We recapitulate known T2D genes such as *HHEX, PAM*, and *CKDN1C*, as well genes where the relevant tissue of action is not whole blood but pancreas or liver (*HNF1A, KCNJ11*). Among the 4 protein associations, gene expression of PAM and ST6GAL1 was also associated with T2D, suggesting a putative continuity from genetic variants to phenotype. Compared to the GTEx v8 results, we find more associations using the larger reference dataset (~0.95% significant associations in DIRECT and from 0.07% to 0.55% in GTEx, depending on tissue). We also find DIRECT implicated genes to be enriched around multiple GWAS loci compared to GTEx tissue derived genes, (odds ratio from 2.2 to 3 in a European GWAS and from 1.4 to 2 in an East Asian GWAS), demonstrating that a higher proportion of DIRECT genes have supporting genetic evidence.

Conclusion

Our findings indicate that large studies can improve the identification of disease causal genes even if a non-relevant tissue is used as a reference panel.

PrgmNr 2399 - Evaluation of genetic prediction and differential expression of genes in non-alcoholic fatty liver disease (NAFLD)

View session detail

Author Block: **T. Kolvekar**[1], W. Li[2], P. Deloukas[1,3], W. Alazawi[2], S. Kanoni[1,3]; [1]Clinical Pharmacology, William Harvey Res. Inst., Barts and the London Med. Sch., Queen Mary Univ. of London, London, United Kingdom, [2]Liver Unit, Blizard Inst., Queen Mary Univ. of London, London, United Kingdom, [3]Ctr. for Genomic Hlth., Queen Mary Univ. of London, London, United Kingdom
Disclosure Block: **T. Kolvekar:** None.
NAFLD is the most common cause of chronic liver disease associated with obesity and metabolic syndrome. It is a disease spectrum ranging from steatosis to non-alcoholic steatohepatitis (NASH) that can progress without symptoms towards end stage liver disease and liver cancer.
Over 11 genetic loci have been associated with NAFLD so far, explaining 10-20% of the heritability. We selected 5,363 patients with NAFLD and 383,165 controls from UK Biobank, all of European ancestry and performed a genome-wide association study (GWAS) analysis. We meta-analysed results with two publicly available summary statistics datasets (1,483 NAFLD cases / 17,781 controls from the EPoS Consortium and 1,106 cases / 8,571 controls from the eMERGE Network). We detected 7 known genetic loci associated with NAFLD at genome-wide significance level and a novel locus, *APOE* (rs429358-T, OR: 1.17, p=7e-11). This missense *APOE* variant was previously associated with CT-measured liver attenuation at exome-wide significance.
We also computed the Fatty Liver Index (FLI) and Fibrosis-4 (FIB-4) and performed separate GWAS analyses in UK Biobank. We used the summary statistics to compute a weighted polygenic risk score (PRS) for each of these indexes, using an optimised pruning and thresholding approach. The best performing PRS for each index was applied to predict classification among NAFLD cases and controls in UK Biobank. Both FLI and FIB-4 were significantly associated with the disease risk but only the FLI-PRS was significantly associated with NAFLD (OR: 1.002 per SD increase, p= 5.66e-85). All quintiles of the FLI-PRS were significantly associated with the risk for NAFLD, compared to the middle quintile. Individuals at the highest 10% of the FLI-PRS had 1.01 higher risk of having NAFLD compared to those at the lowest 10% of the FLI-PRS (p=8.27e-11). The FLI-PRS might be a useful tool to identify individuals at high risk for NAFLD earlier in life.
Finally, we performed RNA-Seq analysis in matched liver and adipose tissue of 25 patients undergoing bariatric surgery (10 NAFLD cases with NASH, 10 NAFLD cases with no NASH and 5 obese controls with normal liver). We identified several differentially expressed genes among the disease states and tissue types implicated in liver injury, fibrosis, inflammation, and metabolism. We found more differentially expressed genes between cases and controls in the adipose tissue, compared to the liver tissue. Simultaneous transcriptomic analysis of liver and fat in NAFLD fibrosis might help disentangle the complex pathophysiology of the disease.

PrgmNr 2400 - Gene-based analyses in population based data identify *MTTP* as associated with nonalcoholic fatty liver disease

View session detail

Author Block: **X. Du**, L. Stetson, Y. Chen, A. Kuppa, GOLD Consortium Groups, E. K. Speliotes; Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **X. Du:** None.
Introduction: Nonalcoholic fatty liver disease (NAFLD) is caused by excess fat (steatosis) deposition in liver, is the most common liver disease worldwide, and has no effective treatments. Hepatic steatosis can be measured using Magnetic Resonance Imaging or Computed Tomography (CT) which correlates with the histologic presence of steatosis with a correlation of 0.78-0.92. NAFLD is heritable/genetic. We and others have identified multiple common single genetic variants that associate with NAFLD; rare variants with effects on NAFLD may be missed by such single variant analyses as they are underpowered. Gene-based analyses which aggregate the effects of rare variants in a gene help to amplify rare variant association signals and can identify genes that when perturbed have effects on NAFLD.
Methods: We used SeqMeta to carry out gene specific association analyses for CT measured liver attenuation which correlates with the presence of liver steatosis in 24,450 multiethnic individuals from the GOLD Consortium. We created SeqMeta objects of 200 KB widths with 100 KB of overlap with neighboring windows and carried out association analysis with inverse normal transformed liver attenuation controlled for age, age2, PCs 1 - 10, alcohol intake as well as kinship structure for family based studies. We filtered for missense and stop-gain variant and performed meta-SKATO analysis. We carried out a similar gene based analysis in UK Biobank where we had liver steatosis measured using magnetic resonance imaging.
Results: In GOLD, *MTTP* (p-value = 2.30E-3) had significant (p-value Conclusions: We can identify genes in population-based studies through gene-based analyses (*MTTP*) that are not captured in single variant analyses. *MTTP* transfers phospholipids and triacylglycerols to nascent apoB for the assembly of lipoproteins. Severe mutations in *MTTP* cause the Mendelian disease abetalipoproteinemia which causes malabsorption in the digestive track and NAFLD due to problems in packaging lipids to be secreted from intestinal or liver cells. Therefore, gene based analyses within large population based studies can now identify bonafide disease genes with rare coding mutations that previously would have been only identifiable with family based analyses, mapping and sequencing.

PrgmNr 2401 - Genetics of osteopontin in patients with chronic kidney disease: the German Chronic Kidney Disease study

View session detail

Author Block: **U. Schultheiss**; Inst. of Genetic Epidemiology, Freiburg, Germany
Disclosure Block: **U. Schultheiss:** None.
Background Osteopontin (OPN), encoded by the *SPP1* gene, is a phosphorylated glycoprotein synthesized predominantly in kidney tissue. OPN levels are known to be associated with adverse kidney outcomes, but the genetic underpinnings of this kidney-enhanced protein are not fully understood. We therefore conducted a genome-wide association study (GWAS) of OPN in a chronic kidney disease (CKD) population, a setting in which the transcription of kidney-specific genes may be altered in comparison to the general population. **Methods** Using genetic data and serum OPN levels quantified via ELISA from CKD patients of European ancestry enrolled into the German Chronic Kidney Disease (GCKD) study, we performed a GWAS of common variants (minor allele frequency [MAF]>1%, additive model), and conducted aggregated rare variant tests (MAFResults Within the GCKD study, data of 4897 participants with a mean age of 60 years (SD 12), median eGFR of 46 mL/min/1.73mÂ² (p25: 37, p75: 57) and median UACR of 50 mg/g (p25: 9, p75: 383) was available. Common variant GWAS revealed index SNPs at 3 loci, two of which replicated in YFS: rs10011284, located downstream of *SPP1*, and rs4253311, mapping into *KLKB1* encoding the protein prekallikrein. The *SPP1* gene was also identified by aggregated rare variant analysis (SKAT p=2.5E-8), comprising 7 variants. The splice acceptor variant (rs139555315) drove the association, with other variants being missense. Using GTEx tissue expression data, colocalization of the association signal with OPN at *SPP1* with pancreas tissue was detected amongst others. SNPs at *KLKB1* showed colocalization of association statistics between serum OPN and various plasma proteins such as SPARC-like and MMP-2 in *trans*, as well as with various phenotypes such as bone disorder and deep venous thrombosis. **Conclusion** This GWAS of OPN levels revealed two replicated associations. The *KLKB1* locus connects the function of OPN with prekallikrein, suggestive of inflammatory functions of the kallikrein-kinin system and theoretically kidney fibrosis. Further studies are needed to elucidated the complexity of OPNâ□□s role within human (patho)physiology.

PrgmNr 2403 - Mirror effects of *OPRD1* variants on diabetes and obesity

View session detail

Author Block: **S. Meulebrouck**[1,2], G. QuÃ©niat[1,2], M. Baron[1,2], M. Canouil[1,2], B. Balkau[3], G. Charpentier[4], S. Franc[4,5], M. Marre[6], R. Roussel[6,7,8], R. Scharfmann[9], A. Bonnefond[1,2,10], P. Froguel[1,2,10]; [1]Inserm UMR1283, CNRS UMR8199, European Genomic Inst. for Diabetes (EGID), Inst. Pasteur de Lille, Lille Univ. Hosp., Lille, France, [2]UniversitÃ© de Lille, Lille, France, [3]Paris-Saclay Univ., Paris-Sud Univ., UVSQ, Ctr. for Res. in Epidemiology and Population Hlth., Inserm U1018, Clinical Epidemiology, Villejuif, France, [4]CERITD (Ctr. dâ□□Ã□tude et de Recherche pour lâ□□Intensification du Traitement du DiabÃ¨te), Evry, France, [5]Dept. of Diabetes, Sud-Francilien Hosp., Univ. Paris-Sud, Corbeil-Essonnes, France, [6]Inserm U1138, Ctr. de Recherche des Cordeliers, Paris, France, [7]Dept. of Diabetology Endocrinology Nutrition, HÃ´pital Bichat, DHU FIRE, Assistance Publique HÃ´pitaux de Paris, Paris, France, [8]UFR de MÃ©decine, Univ. Paris Diderot, Sorbonne Paris CitÃ©, Paris, France, [9]Inst. Cochin, INSERM, U1016, CNRS, UMR8104, UniversitÃ© de Paris, Paris, France, [10]Dept. of Metabolism, Imperial Coll. London, London, United Kingdom

Disclosure Block: **S. Meulebrouck:** None.

Opioid consumption leads to contradictory effects on metabolic homeostasis, by increasing hyperglycemia but improving lipid profile and adiposity. However, the mechanisms linking opioids and metabolism are unknown. RNA sequencing data showed that *OPRD1* encoding Î´ opioid receptor (DOP) is expressed in human islets, especially in Ã□ cells. DOP is an inhibitory G-protein coupled receptor. Based on *OPRD1* resequencing and functional experiments, we aimed to decipher the putative link between *OPRD1* mutations and metabolic disorders.

OPRD1 was sequenced in 6,971 individuals. The effect of each detected variant was assessed via luciferase experiments, in response to increasing concentrations of two DOP agonists (DII or DPDPE). We categorized these variants as gain-of-function (GoF) or loss-of-function (LoF). In parallel, expression and localization of each variant were assessed by western blotting and immunofluorescence assays. Association studies between GoF or LoF variants and various metabolic traits were assessed in our cohort, and in further 34,812 individuals from the T2D Knowledge Portal for the study of the frequent GoF variant encoding p.I52V. Finally, we performed glucose-stimulated insulin secretion (GSIS) in the human Ã□-cell model EndoCÃ□H1 overexpressing *OPRD1* and treated with DII DOP agonist.

In 6,971 individuals, we detected 31 rare variants and 3 frequent variants of *OPRD1*. Luciferase assays highlighted 7 GoF variants, including the frequent variant encoding p.I52V, and 12 LoF variants. Immunofluorescence assays showed that all the mutants were effectively expressed and localized at the plasma membrane, except for two LoF mutants (p.P14R and p.G36E). Western blots showed that these two mutants tended to have a lower expression than wild-type DOP. Association analyses revealed that rare LoF variants increased overweight and obesity risk (P=0.0054; OR=11) but decreased hyperglycemia risk (P=0.054; OR=0.23), while rare GoF variants significantly improved lipid profile. Besides, the frequent GoF p.I52V variant was associated with increased type 2 diabetes risk ($P=3.6Ã□10^{-6}$; OR=2) but decreased body mass index (P=0.0038; Ã□=-0.37) and improved lipid profile, confirming this mirror effect. Finally, we showed that DOP significantly inhibited insulin secretion from Ã□ cells.

This study highlights DOP as a major link between opioids and metabolic disorders. DOP agonists and/or antagonists should be considered as new tools to improve metabolic homeostasis.

PrgmNr 2404 - Mutation Burden Analysis Reveal Genetic Differences between Obesity and non-Obesity Asthma patients in African American Population

View session detail

Author Block: **Y. Liu**[1], J. Qu[2], H-Q. Qu[3], X. Chang[4], F. D. Mentch[5], K. Nguyen[6], J. Glessner[7], P. M. Sleiman[4], H. Hakonarson[8]; [1]Ctr. for Applied Genomics, Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Ctr. for Applied Genomics, Children S Hosp. of Philadelphia, Philadelphia, PA, [3]Children's Hosp. of Philadelphia, Philadelphia, PA, [4]CHOP, Philadelphia, PA, [5]Childrens Hosp. of Philadelphia, Philadelphia, PA, [6]Children's Hosp. of Philadelphia, Phila, PA, [7]Children s Hosp. of Philadelphia, Philadelphia, PA, [8]Children S Hosp. of Philadelphia, Philadelphia, PA

Disclosure Block: **Y. Liu:** None.

Asthma is a complex condition largely attributed to the interactions among genes and environments with a heterogeneous phenotype. Hundreds of asthma-associated genes have been reported, but insufficient replication have made conclusions difficult to draw. Obesity is significantly associated with asthma development, and obesity asthma subjects have more frequent/severe exacerbations while reduced response to asthma medications. In other hands, genetic studies for obese asthma were lacks due to the nature of the complex syndrome. To explore the genetics and potential drug target genes for asthma, especially obese asthma in minority population, we applied blood-derived whole genome sequencing (WGS) data from 4289 African American (AA) individuals. The data include 2226 diagnosed asthma patients, and 1364 subjects who also diagnosed with obesity. The mutation burden analysis was applied using deleterious genomic variants for asthma versus controls, and obese asthma versus non-obese asthma, respectively, based on multiple computational prediction tools. Previously reported asthma associated genes, include *TSLP, RAB6B, SF3B4, RGS7* were identified in asthma/control analysis. For obese asthma comparisons, two gene modules were identified either with significantly adjust p value and/or exclusively occurrences in obese asthma. First group is a list of olfactory receptors genes, second group is immune system related genes that enriched in HDACs deacetylate histones pathway (adjust p value = 7.97E-06), Wnt signaling pathway (5.65E-05), RUNX1 transcription factors pathway (8.22E-05), HATs acetylate histones pathway (1.44E-04), and genomic locus chr6p22.2, which highlighted as contribution of immune system pathways to the genetic architecture of asthma in the most recent genome-wide analysis. As a result, the genes with significantly mutation burden would be the potential drug targets for obese asthma in AA.

PrgmNr 2405 - N6-methyladenosine methylation regulatory genes modified the effect of APOL1 risk genotype on progression of chronic kidney disease

View session detail

Author Block: **C. Li**[1], A. Westbrook[1], R. Zhang[1], L. Hamm[1], J. Chen[1], J. He[1], T. N. Kelly[2]; [1]Tulane Univ. Sch. of Publ. Hlth.and Tropical Med., New Orleans, LA, [2]Tulane Univ, New Orleans, LA
Disclosure Block: **C. Li:** None.
Background: The *APOL1* risk genotype is a major cause of chronic kidney disease (CKD) and CKD progression among African Americans. N(6)-methyladenosine (m6A) is the most abundant modification in mammalian mRNA and regulates all stages of the RNA life cycle. M6A is modified by 10 writer genes which encode m6A methyltransferase enzymes, 9 reader genes which encode m6A binding proteins to stabilize, splice and translate mRNA, and 2 eraser genes encoding m6A demethylase enzymes. We aimed to evaluate whether expression quantitative trait loci (eQTLs) of m6A regulatory genes modify the effect of *APOL1* risk alleles on CKD progression.
Methods: A total of 666 independent (r^2â¤0.3) eQTLs of m6A regulatory genes (Writer genes: *METTL3*, *METTL14*, *WTAP*, *VIRMA*, *HAKAI*, *ZC3H13*, *RBM15*, *RBM15B*, *METTL16*, *PCIF1*/*CAPAM*; Reader genes: *YTHDC1*, *YTHDF1*, *YTHDF2*, *YTHDF3*, *YTHDC2*, *HNRNPA2B1*, *FMR1*/*FMRP*, *HNRNPC*, and *HNRNPG*; and Eraser genes: *FTO* and *ALKBH5*) were retrieved from the GTEx portal. The variants were tested for interactions with *APOL1* risk allele status on CKD progression among 1,224 African American participants of the Chronic Renal Insufficiency Cohort (CRIC) in Cox regression models. We adjusted for age, sex, study site, and the first 10 ancestry principal components (PCs) in a base model and additionally adjusted for baseline kidney function in a full model. We also performed stratified analyses to evaluate *APOL1* risk allele effects on CKD progression according to genotypes of the significant variants.
Results: After Bonferroni correction, a cluster of four variants at the *LAMTOR5* locus (lead SNP: rs6671673, P=6.58Ã10^{-5}) modified the effect of *APOL1* risk alleles on CKD progression. The rs6671673 variant is an eQTL for the writer gene *RMB15*, with the major T allele increasing expression of *RMB15* in whole blood (P=5.4Ã10^{-5}). In stratified analyses, effects of *APOL1* risk alleles on CKD progression decreased with the number of T alleles of rs6671673. Hazard ratios and the corresponding 95% confidence intervals associated with the *APOL1* risk alleles were 1.87 (1.18-2.97), 1.39 (1.13-1.71) and 0.85 (0.67-1.07), respectively, for participants carrying 0, 1, and 2 T alleles at rs6671673.
Conclusion: An eQTL for the m6A writer gene *RMB15* reduced the detrimental effects of the APOL1 risk alleles on CKD progression.

PrgmNr 2406 - Random glucose GWAS in 493,036 individuals provides evidence for reduced lung function as a novel complication of type 2 diabetes

View session detail

Author Block: **A. Ulrich**[1], V. Lagou[2,3,4], L. Jiang[5,6], Z. Balkhiyarova[1,6,7], B. Jones[8], M. Kaakinen[1,6], I. Prokopenko[1,7,9], the Meta-Analysis of Glucose and Insulin-related Traits Consortium (MAGIC); [1]Univ. of Surrey, Guildford, United Kingdom, [2]Wellcome Ctr. for Human Genetics, Oxford, United Kingdom, [3]Human Genetics, Wellcome Sanger Inst., Hinxton, Cambridgeshire, United Kingdom, [4]VIB-KU Leuven Ctr. for Brain and Disease Res., Leuven, Belgium, [5]Inst. for Molecular BioSci., The Univ. of Queensland, Brisbane, Australia, [6]Dept. of Metabolism, Digestion and Reproduction, Imperial Coll. London, London, United Kingdom, [7]Inst. of Biochemistry and Genetics, Ufa Federal Res. Ctr. Russian Academy of Sci., Ufa, Russian Federation, [8]Section of Endocrinology and Investigative Med., Imperial Coll. London, London, United Kingdom, [9]Dept. of Biostatistics, Univ. of Lille, Lille, France
Disclosure Block: **A. Ulrich:** None.

Introduction: Blood glucose and T2D are highly heritable traits (hÂ²â□□30%-60%) with large genome-wide association studies (GWAS) reporting hundreds of associated variants for T2D and over one hundred for fasting glucose. Blood glucose is frequently measured at different times throughout the day (random glucose; RG). Through meta-analysis of RG variant effects in 493,036 non-diabetic individuals from 17 studies, we have expanded by 58 the number of loci associated with glycaemic traits. Previous observational studies have highlighted worsening lung function, as defined by forced vital capacity (FVC), in T2D patients. More recently, it was shown that patients with diabetes are at an increased risk of death from the viral infection COVID-19, with pulmonary dysfunction contributing to mortality. However, confounding and reverse causation inherent to the design of observational studies makes causal inference difficult. Materials and Methods: To quantify the shared genetic contribution between RG and other phenotypes, we estimated their genome-wide genetic correlations using linkage-disequilibrium score regression analyses. We tested causality between RG and correlated traits via bi-directional Mendelian Randomization (MR) analyses using summary level data from non-overlapping samples. Horizontal pleiotropy was tested using the Egger intercept test while heterogeneity was assessed by the Q-statistic. Results: We detected positive genetic correlations between RG, squamous cell lung cancer (rg=0.28, P=0.0015), and lung cancer (rg=0.12, P=0.037); as well as inverse genetic correlations with lung function related traits, such as forced vital capacity (FVC, rg=-0.090, P=0.0059) and forced expiratory volume in 1 second (FEV1, rg=-0.054, P=0.017).MR suggested a causal effect of RG on lung function, including FEV1 ($Î²_{MR-RG}$=-0.60, P=0.0015) and FVC ($Î²_{MR-RG}$=-0.61, P=$3.5x10^{-4}$), but not vice versa. In agreement with this, MR analysis using 413 T2D susceptibility variants suggested T2D as a possible causal factor for declining lung function (FEV1: $Î²_{MR-T2D}$=-0.049, P=$1.27x10^{-13}$; FVC: $Î²_{MR-T2D}$=-0.062, P=$1.42x10^{-21}$). We found no evidence for marked horizontal pleiotropy for the RG-FVC and RG-FEV1 or the T2D-FEV1 MR analyses. We observed no heterogeneity between individual causal estimates for the RG-FVC test (P_{Q-stat}=0.053), while there was some heterogeneity observed for RG-FEV1 (P_{Q-stat}=$6.6x10^{-4}$), and marked heterogeneity for T2D-FVC (P_{Q-stat}=$4.2x10^{-228}$) and T2D-FEV1 (P_{Q-stat}=$7.7x10^{-240}$). Conclusions: Our data suggest a causal effect of glycaemic dysregulation on a decline in lung function as a novel complication of diabetes.

PrgmNr 2407 - The association of fasting glucose with sweetened beverages consumption is greater in Latin Americans with a high polygenic risk score for type 2 diabetes mellitus

View session detail

Author Block: **R. A. Verdugo**[1,2], L. LÃ³pez-Portillo[1], A. Huidobro[3], E. Tobar[1], C. YÃiÃ±ez[1], R. Retamales[1], L. Cifuentes[1], M. L. Moraga[1], M. L. Bustamante[1], P. GonzÃilez-Hormazabal[1], C. Araneda[4,5], V. Maracaja-Coutinho[6,7], C. Ferreccio[8,7]; [1]Programa de Genetica Humana ICBM, Univ. de Chile, Santiago, RM, Chile, [2]Departamento de OncologÃa BÃisico-ClÃnica, Facultad de Med., Univ. de Chile, Santiago, Chile, [3]Departamento de Ciencias PreclÃnicas, Facultad de Med., Univ. CatÃ³lica del Maule, Talca, Chile, [4]Food Quality Res. Ctr., Univ. de Chile, Santiago, RM, Chile, [5]Laboratorio de GenÃ©tica y BiotecnologÃa en Acuicultura, Departamento de ProducciÃ³n Animal, Facultad de Ciencias AgronÃ³micas, Univ. de Chile, Santiago, Chile, [6]Departamento de BioquÃmica y BiologÃa Molecular, Univ. de Chile, Santiago, RM, Chile, [7]Advanced Ctr. for Chronic Diseases, ACCDiS, Santiago, Chile, [8]Sch. of Med., Pontificia Univ. CatÃ³lica de Chile, Santiago, RM, Chile

Disclosure Block: **R.A. Verdugo:** None.

Background and objectives. While sugary drinks intake is decreasing in rich countries, the opposite trend is seen in developing countries; currently the Chilean population presents the highest per-capita consumption world-wide. Consumption of sugar-sweetened beverages alters fasting glucose and other cardiometabolic risk factors. However, it is unclear what effects these lifestyle trend will have among populations of diverse genetic composition. Our goal was to investigate an association between consumption of sugar-sweetened beverages and genetic susceptibility to Type 2 Diabetes (T2D) on fasting glucose in a Chilean population. **Methods.** We calculated a weighted genetic risk score (w-GRS) based on 16 T2D risk SNPs that where genotyped in 2895 non-diabetic participants of the MAUCO cohort by custom targeted sequencing assays. Sugar-sweetened beverage consumption (SSB) was measured with a food frequency questionnaire and classified into 4 levels (0, 1, 2, and >2 servings/day). Log fasting glucose was regressed on SSB and w-GRS tertiles while accounting for socio-demography, lifestyle, and Amerindian ancestry. **Results.** We found that glycemia level increased by SSB (Î²=0.007Â±0.002, p= 0.005) and w-GRS (Î²=0.04Â±0.008, p=0.0001). We also found a significant interaction between sugar consumption and the higher tertile of w-GRS (Î²=0.05Â±0.02, pTCF7L2 (rs7903146, Î²=0.16Â±0.05, pJAZF1 (rs849135, Î²=0.05Â±0.02, pDUSP9 (rs5945326, Î²=0.07Â±0.02, pConclusions. We conclude that association between SSB intake and fasting glucose in a Chilean population is modified by genetic susceptibility to T2D in a non-obesity-mediated manner.

PrgmNr 2408 - The genetic architecture of BMI in infancy and early childhood is characterized by age-specific effects and overlap with pathways involved in Mendelian obesity

View session detail

Author Block: **S. E. Johansson**[1], Ã. Helgeland[2], P. Sole-Navais[3], C. Flatley[4], B. Jacobsson[5], M. Vaudel[6], P. R. Njolstad[7]; [1]Univ. of Bergen, Bergen, Norway, [2]Univ. of Bergen, Oslo, Norway, [3]MÃ¶lndal, Sweden, [4]Norwegian Inst. of Publ. Hlth., Oslo, Norway, [5]Sahlgrenska Univ Hosp, Gothenburg, Sweden, [6]Univ Bergen, Bergen, Norway, [7]Haukeland Univ Hosp., Bergen, Norway
Disclosure Block: **S.E. Johansson:** None.
We performed genome-wide association studies across 12 time points on BMI from birth to eight years of age in 28,681 children and their parents (27,088 mothers and 26,239 fathers) in the Norwegian Mother, Father and Child Cohort Study (MoBa). We identify 46 distinct loci associated with infant and early childhood BMI at specific ages of which 30 have not previously implicated in birthweight or adult BMI. The loci show striking age-dependency in effect sizes that we can model into four main trajectory clusters that align well with the known growth patterns of infancy and early childhood. A total of 21 loci show peak effect between six months and three years, making these discoverable only at early age. Several of the variants reside in/near genes previously implicated in severe forms of early-onset obesity, and monogenic obesity genes are enriched in the vicinity of the 46 loci. Four loci demonstrate evidence of several independent association signals for BMI development near *LEPR*, *GLP1R*, *PCSK1*, and *KLF14*, all central to appetite and energy balance. At the *KLF14* locus, we detect evidence of two distinct regions with imprinting effects. In conclusion, our results provide the most comprehensive map of common genetic influences on early growth to date.

PrgmNr 2410 - Whole-exome sequencing study identifies rare and low-frequency coding variants associated with diabetic nephropathy

View session detail

Author Block: **Y. Pan**[1], X. Sun[1], X. Mi[1], Z. Huang[1], Y. Hsu[2], J. E. Hixson[3], D. Munzy[4], G. A. Metcalf[5], N. Franceschini[6], A. Tin[7], NHLBI Trans-Omics for Precision Medicine (TOPMed)Consortium, TOPMed Kidney disease working group, M. Fornage[8], Y-D. I. Chen[9], A. Correa[10], R. S. Vasan[11], D. K. Arnett[12], B. D. Mitchell[13], C. Kooperberg[14], R. P. Tracy[15], J. I. Rotter[16], R. A. Gibbs[5], A. C. Morrison[17], H. I. Feldman[18], E. Boerwinkle[8], J. He[1], T. N. Kelly[19]; [1]Tulane Univ., New Orleans, LA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Univ Texas Sch Publ. Hlth, Houston, TX, [4]Baylor Coll. of Med., Houston, TX, [5]Baylor Coll. Med., Houston, TX, [6]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [7]Jackson, MS, [8]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [9]LOS ANGELES BIOMED Inst., Torrance, CA, [10]Univ. of Mississippi Med. Ctr., Jackson, MS, [11]Boston Univ Sch Med, Boston, MA, [12]Univ Alabama Birmingham, Birmingham, IN, [13]Univ Maryland, Baltimore, Baltimore, MD, [14]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [15]Univ. of Vermont, Burlington, VT, [16]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [17]Univ. of Texas at Houston, Houston, TX, [18]Univ. Of Pennsylvania Sch. Of Med., Philadelphia, PA, [19]Tulane Univ, New Orleans, LA

Disclosure Block: **Y. Pan:** None.

Diabetic nephropathy (DN) is a major public health challenge, posing substantially increased risks of end-stage-renal disease, cardiovascular disease, and death. Although established as a polygenic condition, the genomic mechanisms underlying DN remain largely unknown. To reveal novel genes and variants associated with DN, we conducted a whole-exome sequencing study among -4) from single variant and aggregate rare variant analyses of multi-ancestry and ancestry-specific participants were replicated leveraging whole-genome sequencing data from 11,487 multi-ancestry Trans-Omics for Precision Medicine (TOPMed) participants. In single variant analyses, two novel variants achieved exome-wide significance (P-7) in discovery and replication stage meta-analyses comparing DN cases to healthy controls. A rare inframe insertion (MAF=0.0021) at the novel *DIS3L2* locus was associated with increased risk of DN in multi-ancestry (OR=193, P=3.59Ã□10^{-9}) and white (OR=100, P=3.01Ã□10^{-8}) participants. In addition, a low-frequency variant (MAF=0.038) at the novel *KRT6B* locus strongly associated with DN in whites (OR=5.31, P=2.72Ã□10^{-9}). Furthermore, we identified aggregate associations of rare predicted loss-of-function variants in *ERAP2* (P=4.03Ã□10^{-8}) and *NPEPPS* (P=1.51Ã□10^{-7}) among multi-ancestry participants at exome-wide significance (P-6) in meta-analyses comparing those with DN to healthy controls and kidney disease controls, respectively. In general, reported signals were consistent across analyses of the three control groups. In summary, we identified and replicated novel signals with large effects in exonic regions, providing empirical evidence of rare variants with large effects on DN and highlighting the utility of aggregate tests to complement single marker analyses for rare variant association discoveries.

PrgmNr 2411 - Defining the Regulatory Mechanism for Helios in Autoimmune Disease

View session detail

Author Block: **S. Guga**[1], D. C. Graham[1], T. J. Vyse[2]; [1]King's Coll. London, London, United Kingdom, [2]King S Coll. London, London, United Kingdom
Disclosure Block: **S. Guga:** None.
Systemic Lupus Erythematosus is a complex autoimmune disease of which the mechanism is unclear. Our group has completed a large genome-wide genetic association study in SLE. One of the genes we identified was IKZF2 (Helios) - a member of the Ikaros family of Kruppel zinc finger transcription factors (TF). Studies have implied the important role of IKZF2 (Helios) in the development and function of the immune system. In the current project, we aim to elucidate the function of Helios in different T cell types by defining its binding sites and related target genes and enriched pathways Jurkat T cells, regulatory T cells with combinatory use of several bioinformatics tools multiple large-scale of high throughout sequencing datasets and obtain an implication about the role of IKZF2 in development of SLE. Up until the date of this submission, we explored for the first time the genome-wide binding sites and direct target genes and regulatory network of IKZF2 in Jurkat T cell by combining chip-seq data with gene expression data. The differential expression between Jurkat T cell with a Helios knockdown and wild-type Jurkat T cells were analyzed. 415 genes (115 upregulated and 300 downregulated) 4 were found to be differentially expressed between Helios knockdown Jurkat T cell and wild type Helios Jurkat T cell (with the cutoff of |log2FC|>1.0, adjust P value

PrgmNr 2412 - Evaluate mitophagy/autophagy modulators and FDA-approved drugs to prevent or ameliorate the autoimmune, lipodystrophic and neurodegenerative phenotypes observed in *Clec16a* KO mice

View session detail

Author Block: **R. Pandey**, B. Strenkowski, M. Bakay, H. Hakonarson; Children's Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **R. Pandey:** None.
Abstract: CLEC16A is implicated in multiple autoimmune diseases. We generated an inducible whole-body knockout (KO), Clec16aî UBCâ¯ mice to address the role of CLEC16A loss of function. Preliminary data from our inducible *Clec16a* KO mice suggests a derailed functional link between mitophagy/autophagy/lipophagy, SOCS signaling and CLEC16A, predisposing to the immune dysregulation and the risk of developing resulting in an inflammatory phenotype and/or neurodegeneration. We propose to examine if treatment with drugs targeting mitophagy/autophagy/lipophagy/SOCS will rescue the phenotypic defects and prevent the onset of the autoimmune, inflammatory and neurodegenerative phenotypes observed in CLEC16A KO mice.
Methods: We generated *Clec16a* inducible knockout (KO) mice to address the role of CLEC16A loss of function in autoimmune inflammatory disorders. Turning off *Clec16a* in adult mice leads to dysregulated mitophagy, severe weight loss, robust autoimmune inflammatory responses, severe neurological symptoms with neuroinflammation and progressive neurodegeneration resembling spinocerebellar ataxia. Here, we examined if the observed perturbations in mitophagy/autophagy, cytokine expression, release and signaling, can be rescued by interventions targeting the mitophagy/autophagy with or without anti-inflammatory therapy, using, JAK-STAT pathway inhibitors, such as SOCS1 and tofacitinib. **Results:** Treatment with a JAK/STAT inhibitor (tofacitinib) partially rescued the inflammatory lipodystrophic phenotype and improved survival of *Clec16a* KO mice by modulating ER stress, lipolysis, mitophagy and autophagy. Our results provide evidence in support of phenotype rescue from targeting autophagy/mitophagy and in combination with SOCS1. In patient populations harboring variants that result in CLEC16A hypofunction, drugs with modulatory effects on mitophagy/autophagy in combination with JAK-STAT signaling inhibition, could compensate for the attenuated CLEC16A activity and present formidable candidates for targeted interventions.
Conclusion: Our mouse model serves as a valuable tool for assessing therapeutic interventions that may and will prevent/delay the phenotype occurrence and progression in autoimmune disease patients and improve the outcome in patient populations harboring variants that result in CLEC16A hypofunction.

PrgmNr 2413 - Expanded genome-wide association study of IBD identifies novel loci and implicates new genes in disease susceptibility

View session detail

Author Block: **L. Fachal**, International IBD Genetics Consortium; Wellcome Sanger Inst., Hinxton, United Kingdom
Disclosure Block: **L. Fachal:** None.
Background: Genome-wide association studies (GWASs) have identified 241 loci associated with inflammatory bowel disease (IBD). However, the mapping of additional disease loci and fine-mapping causal variants is still limited by sample size. Larger GWASs can provide further insights into causal IBD biology.
Methods: We performed a GWAS meta-analysis of 33 cohorts, totalling 54,439 IBD patients (30,574 with Crohnâ□□s disease (CD) and 21,193 with Ulcerative Colitis (UC)) and 37,054 European population controls. Genotype imputation was undertaken using the TOPMed diverse population panel and logistic mixed model association tests were performed using REGENIE.
Results: We identified 78 novel genome-wide significant (p1Mb from any known loci. Five new loci contain genes implicated in monogenic autosomal recessive syndromes that include colitis: *CARMIL2*, *DOCK8*, *G6PC3*, *HPS4*, *NCF1*. Many new loci causally alter the expression of nearby genes (ie *DOCK8*, *ITPKC*, *SDCBP2, STK10*) in relevant cell types and tissues, suggesting that aberrant expression of these genes underpins the IBD association. Fine-mapping analyses identified likely causal missense variants at three new loci. *DOK2* (OR_{CD}=1.24 p=7.4x10-9) encodes a cytoplasmic signalling protein highly expressed in macrophages and T cells in the terminal ileum. Loss of Dok2 in mice causes severe DSS-induced colitis with reduced IL17A and IL22 expression. *SHARPIN* (OR_{CD}=1.25 p=6.7x10-17) forms part of the linear ubiquitin chain assembly complex that modulates activation of the NF-ÎºB signalling pathway. Loss-of-function (LOF) mutations in other proteins in the complex are associated with both combined immunodeficiency and systemic autoinflammation. *CARMIL2* (OR_{UC}=1.23 p=1.1x10-9) is required for NF-ÎºB signalling in both B and T cells, and is expressed in these cell populations in the terminal ileum. LOF mutations are associated with primary immunodeficiencies and paediatric forms of IBD.
Discussion: We performed the largest GWAS meta-analysis of IBD to date, which enabled us to identify low frequency variants with larger effects on IBD susceptibility than the more common variants typically identified via GWAS. The biological overlap between Mendelian and more complex forms of IBD is demonstrated by the identification of common non-coding variants associated with complex forms of IBD that dysregulate the function of a Mendelian gene. At the meeting we will present results of our fine-mapping and pathway analyses, with a view to identifying candidate drug targets for IBD.
Wellcome Trust WT206194 NIDDK U24 DK062429

PrgmNr 2414 - Risk HLA alleles in South America and potential new epitopes for SARS-CoV2

View session detail

Author Block: **S. SAENZ**[1], V. I. Romero[2]; [1]UNIVERSIDAD SAN FRANCISCO DE QUITO, QUITO, Ecuador, [2]Univ. San Francisco de Quito, Quito, Ecuador
Disclosure Block: **S. Saenz:** None.
HLA alleles are associated with the bodyâ□□s response to infection, such as SARS-CoV2. We asked which HLA alleles were associated with protection or susceptibility to SARS-CoV2 infection and the prevalence of these HLA alleles in South America. We performed a literature review which found that previous reported HLA protective/susceptible allele, obtained the prevalence of those HLA in South America, as well as the most common HLA alleles in this region. Also, we calculated binding prediction of the most common HLA alleles in this continent to SARS-CoV2 immunogenic regions. We identified that the previously reported protective/susceptible alleles are not frequent in South America, confirmed that the spike protein is the most immunogenic protein of SARS-CoV2, and detected new immunogenic epitopes that bound to protective HLA alleles and to HLA alleles common in South America (binding score>0.90). The epitopes found in our study could be used as vaccine targets. Information related to regional HLA frequency databases in South America is limited; therefore, our study reinforces the importance of researching HLA variants in different populations and encourages the creation of a South America HLA variant database.

PrgmNr 2415 - SARS-CoV-2 as a trigger of autoimmune diseases: brazilian experience

View session detail

Author Block: **M. C. Silva**[1], G. F. S. Carvalho[1], V. C. QuintÃ£o[1], Y. G. Oliveira[1], V. T. Almeida[1], A. M. Nascimento[1], B. M. Wolff[1], L. L. Vieira[1], C. S. Faria[1], L. Antonangelo[1], L. D. Kulikowski[2]; [1]Univ. of SÃ£o Paulo, SÃ£o Paulo, Brazil, [2]Univ. de Sao Paulo, Sao Paulo, Brazil
Disclosure Block: **M.C. Silva:** None.
The infection by the new virus SARS-CoV-2, responsible for causing the coronavirus disease 2019 (COVID-19), has been reported as a trigger for several different autoimmune diseases, which are characterized by an abnormal reaction from the immune system against the body's healthy cells. These diseases have a complex activation mechanism including genetic factors associated with susceptibility. Different genome-wide association studies and meta-analysis identified single nucleotide polymorphisms (SNPs) possibly related, which represents a genetic susceptibility in some individuals to develop autoimmune diseases. In the present study, we purpose to investigate the impact of 10 SNPs, for each disease, who increase the susceptibility for the development of diabetes mellitus type 1, psoriasis, Hashimotoâ□□s thyroiditis, asthma, multiple sclerosis, systemic lupus erythematosus and rheumatoid arthritis, which represents seven of the most important autoimmune diseases previously observed in clinic practice and reported as triggered by the SARS-CoV-2 infection. We evaluated 24 DNA samples from patients previously infected by the SARS-CoV-2 virus using the SNP-array technique with the Infinium ImmunoArray-24 v2 BeadChip Kit. The analyses of the SNPs and genomic regions were obtained from GenomeStudio 2.0 and RStudio softwares and the variants were selected using the following databases: GWAS Catalog, dbSNP, Database of Genomic Variants (DGV) e UCSC Genome Bioinformatics. The results allowed us to identify two patients, who developed multiple sclerosis after the SARS-CoV-2 infection, with allelic compatibility for six of the ten selected genomic variants associated with the disease (rs1077667, rs12927355, rs1323292, rs1738074, rs1800693 e rs2104286). The patient 1 has 4 SNPs in homozygosis for the impact allele, while the patient 2 has two of them in homozygosis. Furthermore, both patients have an allelic homozygosis to the same variant rs2104286 in the *IL2RA* gene, showing a strong association genotype-phenotype. In the current scenario, COVID-19 represents a global challenge for health policies, including many diseases reported in post-COVID. Thus, investigation of specific SNPs can directly contribute to improve the understanding of genomic variants considered as a potential trigger for important autoimmune diseases such as multiple sclerosis.

PrgmNr 2416 - Whole-exome sequencing of hundreds of microbiopsies reveals the somatic evolutionary landscape of psoriatic skin

View session detail

Author Block: **S. Olafsson**[1], E. Rodriguez[2], Y. Hooks[1], I. Martincorena[1], S. Weidinger[2], P. J. Campbell[1], C. A. Anderson[1]; [1]Wellcome Trust Sanger Inst., Hinxton, Cambridgeshire, United Kingdom, [2]Dept. of Dermatology and Allergy, Univ. Hosp. Schleswig-Holstein, Campus Kiel, Kiel, Germany
Disclosure Block: **S. Olafsson:** None.
Recent studies of somatic evolution have revealed that normal tissues are a mosaic of cellular clones competing for limited space within the body. In some instances, clones carrying favorable mutations may transform into cancers. However, it is becoming increasingly clear that somatic evolution may also contribute to non-neoplastic disease, for example cardiovascular disease and inflammatory bowel disease. Here we study somatic evolution in psoriasis, a chronic inflammatory disease of the skin affecting about 2% of Europeans. We have used laser capture microscopy to isolate samples of keratinocytes, 2 in size, from psoriatic lesions and adjacent non-lesional skin. By whole-exome sequencing hundreds of microbiopsies from dozens of donors, we compare the psoriatic and normal epidermis in terms of its clonal structure, mutation burden, mutagen activity and selection landscape. Results from the first 140 exomes indicate that in the absence of cytotoxic treatment, the clonal structure of psoriatic skin is similar to that of non-lesional skin, with clones rarely expanding over large distances. As in non-lesional skin, most mutations in psoriatic skin result from UV-light exposure, but we reveal a large heterogeneity in UV-associated mutation burden between cells separated by less than 1mm in the tissue. We identify genes under selection in the skin and compare their mutational frequencies in diseased and healthy tissue. Preliminary results replicate earlier findings of frequent mutations in *NOTCH1*, *FAT1* and *PPM1D* in the skin and additionally suggest positive selection of mutations in *GXYLT1*. The encoded protein is involved in the glycosylation of epidermal growth factor motifs of proteins such as NOTCH1 and the mutations likely represent another mechanism for disrupting NOTCH signaling. Finally, we describe one patient with a history of phototreatment with psoralens. This cytotoxic treatment left a characteristic mutational signature, caused hundreds of thousands of mutations in the genomes of exposed cells and enabled the expansion of heavily mutated clones in the skin. Results from an expanded cohort of over 1000 exomes will be presented at the meeting.

PrgmNr 2417 - A *de novo* paradigm for male infertility

View session detail

Author Block: **M. Xavier**[1], M. S. Oud[2], R. Smits[3], H. E. Smith[1], F. K. Mastrorosa[1], G. S. Holt[1], B. J. Houston[4], P. F. de Vries[2], B. K. S. Alobaidi[1], L. E. Batty[1], H. Ismail[1], J. Greenwood[5], H. Sheth[6], A. Mikulasova[1], G. Astuti[7], C. Gilissen[8], K. McEleny[9], H. Turner[10], J. Coxhead[11], S. Cockell[12], D. Braat[13], K. Fleischer[3,14], K. Dâ□□Hauwers[15], E. Schaafsma[16], GEMINI Consortium, L. Nagirnaja[17], D. F. Conrad[17], C. Friedrich[18], S. Kliesch[19], K. I. Aston[20], A. Riera-Escamilla[21], C. Krausz[22], C. G. Gonzaga-Jauregui[23], M. F. Santibanez-Koref[24], D. J. Elliott[1], L. E. Vissers[25], F. TÃ¼ttelmann[18], M. O'Bryan[26], L. Ramos[3], G. W. van der Heijden[2,3], J. A. Veltman[27]; [1]BioSci.s Inst., Faculty of Med. Sci., Newcastle Univ., Newcastle upon Tyne, United Kingdom, [2]Dept. of Human Genetics, Donders Inst. for Brain, Cognition and Behaviour, Radboudumc, Nijmegen, Netherlands, [3]Dept. of Obstetrics and Gynaecology, Radboudumc, Nijmegen, Netherlands, [4]Sch. of BioSci., Faculty of Sci., The Univ. of Melbourne, Melbourne, Australia, [5]Dept. of Genetic Med., The Newcastle upon Tyne Hosp. NHS Fndn. Trust, Newcastle upon Tyne, United Kingdom, [6]FRIGE's Inst. of Human Genetics, Ahmedabad, India, [7]Dept. of Human Genetics, Radboud Inst. for Molecular Life Sci., Radboudumc, Nijmegen, Netherlands, [8]Radboud Univ. Med. Ctr., Nijmegen, Nederland, Netherlands, [9]Newcastle Fertility Ctr., The Newcastle upon Tyne Hosp. NHS Fndn. Trust, Newcastle upon Tyne, United Kingdom, [10]Dept. of Cellular Pathology, The Newcastle upon Tyne Hosp. NHS Fndn. Trust, Newcastle upon Tyne, United Kingdom, [11]Genomics Core Facility, Faculty of Med. Sci., Newcastle Univ., Newcastle upon Tyne, United Kingdom, [12]Bioinformatics Support Unit, Faculty of Med. Sci., Newcastle Univ., Newcastle upon Tyne, United Kingdom, [13]Dept. of Obstetrics and Gynaecology, Radboudumc, Newcastle upon Tyne, United Kingdom, [14]TFP Ctr. of Reproductive Med., DÃ¼sseldorf, Germany, [15]Dept. of Urology, Radboudumc, Nijmegen, Netherlands, [16]Dept. of Pathology, Radboudumc, Nijmegen, Netherlands, [17]Div. of Genetics, Oregon Natl. Primate Res. Ctr., Oregon Hlth.& Sci. Univ., Beaverton, OR, [18]Inst. of Reproductive Genetics, Univ. of MÃ¼nster, MÃ¼nster, Germany, [19]Ctr. of Reproductive Med. and Andrology, Dept. of Clinical and Surgical Andrology, Univ. Hosp. MÃ¼nster, MÃ¼nster,, Germany, [20]Univ of Utah, Salt Lake City, UT, [21]Andrology Dept., FundaciÃ³ Puigvert, Univ.t AutÃ²noma de Barcelona, Inst. de Investigaciones BiomÃ©dicas Sant Pau (IIB-Sant Pau), Barcelona, Spain, [22]Dept. of BioMed., Experimental and Clinical Sci. â□□Mario Serioâ□□, Univ. of Florence, Florence, Italy, [23]Regeneron Genetics Ctr., Tarrytown, NY, [24]Newcastle Univ., Newcastle upon Tyne, United Kingdom, [25]Radboudumc, Nijmegen, Netherlands, [26]Sch. of BioSci., Faculty of Sci., The Univ. of Melbourne, Parkville, Australia, [27]BioSci.s Inst., Faculty of Med. Sci., Newcastle Univ., Newcastle-upon-Tyne, United Kingdom

Disclosure Block: **M. Xavier:** None.

De novo mutations (DNMs) are known to play a prominent role in many sporadic disorders with reduced fitness and genetic heterogeneity. Due to this strong effect on fitness, we hypothesize that DNMs play a prominent role in male infertility and explain a significant proportion of the genetic causes in this understudied disorder, where large-scale studies have to yet been published. In our study, we performed trio-based exome sequencing in a unique cohort of 185 males with unexplained cases of azoospermia or oligozoospermia and their unaffected parents. In total, 145 rare protein-altering *de novo* SNVs and 2 *de novo* CNVs were identified in these patients. Following a systematic analysis assessing mutational impact and protein function, 29 DNMs were classified as possibly causative of the male infertility phenotype observed in the affected patients. Additionally, a significant enrichment was detected in the number of Loss-of-Function (LoF) DNMs in LoF-intolerant genes ($p=1.00x10^{-5}$) and in predicted pathogenic missense DNMs in missense-intolerant genes ($p=5.01x10^{-4}$). Overall, a significant increase was found in the number of protein-protein interactions amongst genes affected by these DNMs ($p=2.35x10^{-2}$). Among the new candidate genes identified was *RBM5*, an essential regulator of male germ cell pre-mRNA splicing. Besides the patient carrying

the DNM in *RBM5*, 6 additional infertile men were found carrying a distinct rare pathogenic missense mutation in *RBM5* in an international cohort of patients (n=2,506), a significant enrichment when compared to the number of mutations found in *RBM5* in a cohort of confirmed fertile men (n=5,784; p=0.03).Taken together, our results provide strong evidence for the role of DNMs in severe male infertility and identify a number of new candidate genes affecting human male fertility.

PrgmNr 2418 - Analysis of transcriptional changes associated with pubertal development

View session detail

Author Block: **J. Resztak**[1], J. Choe[2], R. Bruinsma[1], R. Houpt[3], R. B. Slatcher[4], S. Zilioli[1], R. Pique-Regi[5], F. Luca[5]; [1]Wayne State Univ., Detroit, MI, [2]New York Univ., New York, NY, [3]Arizona State Univ., Tempe, AZ, [4]Univ. of Georgia, Athens, GA, [5]Wayne State Univ, Detroit, MI
Disclosure Block: **J. Resztak:** None.
Puberty is an important developmental period marked by hormonal, metabolic, immune and physiological changes, which have been implicated in predisposition to immune disease later in life. Yet little is known about the immune gene expression changes accompanying pubertal development. We studied a cohort of 251 children (103 girls, 148 boys, 10-17 years old), assessing their pubertal development and leukocyte gene expression changes between two time points (1-2 year interval). We identified gene expression changes over this time interval for 3240 genes in boys and 3187 genes in girls, which were largely overlapping (Pearsonâ□□s correlation=0.9, p-valueDSC1 and *TRBV30* in females and *HLA-H* and *PGAP1* in males (10% FDR). *HLA-H* was previously associated with auto-immune disease risk, including rheumatoid arthritis and multiple sclerosis in transcriptome-wide association studies. Overall, we demonstrated that changes in leukocyte gene expression reflect the physiological changes associated with pubertal development. Insight into immune gene expression processes accompanying puberty can help us understand their contributions to immune disease later in life.

PrgmNr 2419 - Construction of copy number variation map identifies small regions of overlap and candidate genes for atypical female genitalia

View session detail

Author Block: **A. Amukamara**[1], I. E. Amarillo[2]; [1]Washington Univ. Sch. of Med., St. Louis, MO, [2]Washington Univ in St Louis Sch. of Med., St. Louis, MO
Disclosure Block: **A. Amukamara:** None.
Copy number variations (CNVs) have been implicated in various conditions of differences of sexual development (DSD). Generally, larger genomic aberrations are more often considered disease-causing or clinically relevant, but over time, smaller CNVs have been associated with various forms of DSD. The main objective of this study is to identify small CNVs and smallest regions of overlap (SROs) in patients with atypical female genitalia (AFG) and build a CNV map of AFG. We queried the DECIPHER database for recurrent duplications and/or deletions detected across the genome of AFG individuals. From these data, we constructed a chromosome map consisting of SROs and investigated such regions for genes that may be associated with the development of atypical female genitalia. Our study identified 180 unique SROs (7.95-45,338.26 kb) distributed among 22 chromosomes. The most SROs were found in chromosomes X, 17, 11, and 22. None were found in chromosome 3. From these SROs, we identified 22 genes as potential candidates. Although none of these genes are currently associated with AFG, literature review indicated that almost half were potentially involved in the development and/or function of the reproductive system, and only one gene was associated with a disorder that reported an individual patient with ambiguous genitalia. Our study further demonstrates the clinical significance of small CNVs, and additional downstream functional investigations will provide better understanding of the genetic etiology of AFG. Data from this investigation may also aid in genetic counseling, genetic diagnosis, and management of patients with AFG.

PrgmNr 2420 - Genetic architecture of severe spermatogenic failure and male infertility: time to advance standard of care

View session detail

Author Block: **C. Friedrich**[1], M. J. Wyrwoll[1], A-K. Dicke[1], E. Pohl[1], N. KÃ¶ckerling[1], N. Rotte[1], J. Emich[1], M. Vockel[2], M. WÃ¶ste[3], C. Ruckert[2], J. Seggewiss[2], S. Ledig[2], A-C. Tewes[2], Y. Stratis[2], J. Wistuba[4], C. Krallmann[4], S. Kliesch[4], A. RÃ¶pke[2], B. Stallmeyer[1], F. TÃ¼ttelmann[1]; [1]Inst. of Reproductive Genetics, Univ. of MÃ¼nster, MÃ¼nster, Germany, MÃ¼nster, Germany, [2]Inst. of Human Genetics, Univ. of MÃ¼nster, MÃ¼nster, Germany, MÃ¼nster, Germany, [3]Inst. of Med. Informatics, Univ. Hosp. MÃ¼nster, MÃ¼nster, Germany, MÃ¼nster, Germany, [4]Ctr. of Reproductive Med. and Andrology, Univ. Hosp. MÃ¼nster, MÃ¼nster, Germany, MÃ¼nster, Germany
Disclosure Block: **C. Friedrich:** None.

Half of all infertility cases, affecting 10-15% of couples, arise from male factors like crypto- and azoospermia (very few or no sperm in the semen). The only well-established genetic causes include Klinefelter syndrome (KS, karyotype 47,XXY), Y-chromosomal AZF microdeletions for quantitative spermatogenic failure (SPGF), and *CFTR* mutations for obstructive azoospermia (OA). However, although gene mutations are expected to explain most remaining cases, no respective analyses have yet informed clinical practice.

To address this, since 2017, we prospectively included crypto- and azoospermic men attending the Centre of Reproductive Medicine and Andrology (CeRA, MÃ¼nster, Germany) into our Male Reproductive Genomics (MERGE) study. After excluding known conditions leading to SPGF (e.g., testicular tumors and/or previous chemotherapy) and ruling out chromosomal aberrations and AZF deletions, we offered exome sequencing, analyzed 62 genes with at least limited clinical validity to be associated with crypto-/azoospermia based on our recent review of overall 596 genes, and strictly assessed variants according to ACMG-AMP guidelines.

Chromosomal aberrations (mostly KS) and AZF deletions were identified in 15.9% and 2.8% of cases. In 682 men without such conditions, 18 patients (2.6%) with OA carried causal variants in *CFTR* or *ADGRG2*. In 40 patients (5.9%) with SPGF, variants in 21 genes were identified as (likely) pathogenic. The most commonly affected genes were *TEX14* (5 cases), *NR5A1* (4 cases), and *AR*, *M1AP*, *SYCP2* (each 3 cases). Segregation analyses helped in assessing the pathogenicity of some variants, highlighted by a de novo variant in *DMRT1*.

The sequencing performed here offers a substantial diagnostic yield on top of that of current genetic tests. Further, some findings are immediately relevant for infertility counselling/treatment, as they indicate the success of treatments like testicular biopsy with the aim of sperm extraction (TESE). Overall, panel/exome sequencing provides strong enough evidence to be included in diagnostic workups of men with crypto- and azoospermia. Including more genes and additional functional assessments of variants of uncertain significance will quickly increase relevant diagnostic findings and improve patients care.

PrgmNr 2421 - Idiopathic hypogonadotropic hypogonadism in the Maltese island population: A spectrum of phenotypes and genotypes

View session detail

Author Block: **C. Axiak**[1], A. Pleven[1,2], F. Borg Carbott[1], R. Attard[1], K. Cassar[3], M. Gruppetta[3,4], J. Vassallo[3,4,5], S. Bezzina Wettinger[1,5], R. Farrugia[1,5]; [1]Dept. of Applied BioMed. Sci., Faculty of Hlth.Sci., Univ. of Malta, Msida, Malta, [2]Clinical Chemistry Section, Dept. of Pathology, Mater Dei Hosp., Msida, Malta, [3]Dept. of Med., Faculty of Med. and Surgery, Univ. of Malta, Msida, Malta, [4]Div. of Endocrinology and Diabetes, Dept. of Med., Mater Dei Hosp., Msida, Malta, [5]Ctr. for Molecular Med. and Biobanking, Faculty of Med. and Surgery, Univ. of Malta, Msida, Malta

Disclosure Block: **C. Axiak:** None.

Idiopathic Hypogonadotropic Hypogonadism (IHH) is a rare, genetically heterogeneous infertility disorder characterized by absent or incomplete sexual maturation by age 18. This is due to a deficiency or absence of either pulsatile hypothalamic secretion of gonadotropin-releasing hormone (GnRH) and/or gonadotropins from the anterior pituitary in the presence of low serum sex steroids. IHH diagnosis is only conclusive if there are no other abnormalities to hypothalamic and pituitary morphology and function.

Here, we report findings from a targeted gene panel of 17 IHH genes (*CHD7, EBF2, FGF8, FGFR1, GNRH1, GNRHR, HS6ST1, KAL1 (ANOS1), KISS1, KISS1R, NSMF (NELF), PROK2, PROKR2, TAC3, TACR3, TSHZ1,* and *WDR11*) on 15 probands with IHH. Causative variants explaining the phenotype were identified in 4 cases. These comprise (i) oligogenic inheritance of autosomal dominant *FGFR1* p.Y132C and autosomal recessive *TACR3* p.K286R in a male with cryptorchidism and a micropenis, (ii) a novel autosomal dominant *CHD7* p.Q78X stop variant in an anosmic male with 2 CHARGE characteristics and cryptorchidism, (iii) compound heterozygosity for *GNRHR* p.Q106R and the novel *GNRHR* p.N298KfsTer22 in a female, and (iv) homozygosity for a novel autosomal recessive *KISS1R* p.Y190_A199del deletion in a male. Four other probands were found to have only variants with partial or complete loss of function in heterozygosity and are thus only partially solved. Of these, 2 were females who presented with secondary amenorrhea. The first had an accumulation of heterozygous kisspeptin variants including *KISS1R* p.Q36R that is reported to contribute to pathogenicity. The second had an autosomal recessive *GNRHR* p.Q106R heterozygote variant. Exome sequencing on this latter case and her hyposmic father, who was diagnosed with IHH and took human chorionic gonadotropin treatment to achieve fertility, showed that the father was also heterozygous for *GNRHR* p.Q106R and had an additional novel heterozygous *CCDC141* p.E905K variant within a highly conserved region. Pathogenic variants in this gene are suspected to cause decreased embryonic GnRH cell migration that precedes formation of the hypothalamic neuronal network that secrets GnRH.

We also found *GNRHR* p.Q106R in heterozygosity to be a recurring variant in 4 unrelated individuals from the cohort. The local allele frequency was found to be 0.029 in a population study using 493 cord blood DNA samples of neonates with at least 1 ethnically Maltese parent. For a variant that is reported to contribute to IHH pathogenicity in multiple literature accounts, the variant frequency in the local population is higher than expected and suggests a founder effect.

PrgmNr 2422 - Challenging diversity in precision medicine through large-scale exome sequencing and genome wide genotyping in South Asians: Analyses of quantitative red blood cell traits

View session detail

Author Block: **S. Tuna**[1], W. McLaren[1], N. England[1], N. S. Roy[2], A. Shah[3], J. L. Rodriguez-Flores[4], M. Kapoor[4], Y. Turpaz[5], S. Jamuar[5], A. R. Schuldiner[4], J. Picker[6]; [1]Global Gene Corp, Cambridge, United Kingdom, [2]Global Gene Corp, Mumbai, India, [3]Global Gene Corp, Ahmedabad, India, [4]Regeneron Pharmaceuticals, Inc., Tarrytown, NY, [5]Global Gene Corp, Singapore, Singapore, [6]Global Gene Corp, Boston, MA
Disclosure Block: **S. Tuna:** None.
Background: Red blood cells (RBC) are essential for oxygenation and when disrupted can cause significant health problems. India has a high prevalence for RBC disorders including thalassemias, sickle cell anemias and others. Most genetic analyses of RBC traits report findings from European populations. As genomic architecture differs among populations, e.g., linkage disequilibrium (LD) structure and allele frequency, this bias can cause studies to miss significant variants that are not seen in European populations.
Methods: With the goal of increasing diversity in genomic studies and to better understand correlation between these markers and RBC disorders, we collected and sequenced (whole exome sequencing and chip array) >15,000 South Asian samples to date and ascertained complete blood count that included standard red blood cell indices (haemoglobin, mean corpuscular volume (MCV), hematocrit). We performed exome and genome wide association analyses of both common and rare (AF **Results**: MCV, a measurement of the size of the RBC, had the most significant results in our study. We identified several known and novel MCV-associated loci specific to South Asian populations, including *CLCN6* (leading variantâ□□s beta = -1.21(0.17), p-value = 1.4e-12), HBB (leading variantâ□□s beta = -2.15(0.26), p-value=3.7e-16), *LUC7L* (leading variantâ□□s beta = -5.07(0.42), p-value = 2.0e-33), and *SCO2* (leading variantâ□□s beta = -0.91(0.14), p-value=1.2e-10). Using a rare variant gene burden test approach we replicated *HBB* (beta = -19.20(1.10), p-value = 1.8e-68) and *SH3PXD2A* (beta = -6.29(1.16), p-value=6.4e-08.
Conclusion: Our study identified novel and known variants associated with RBC traits in South Asians. It is possible that some of these associations reflect adaptation or susceptibility to blood disorders. Better understanding of the correlation between these markers and RBC disorders within the population will lead to improved understanding of disease burden. Furthermore, a better understanding of complex traits and diseases through genetic studies of diverse populations will lead to novel drug targets, biomarker identification for patient stratification, and the promise of precision medicine for all.

PrgmNr 2423 - Clonal myelopoiesis and risk of adverse events in chronic kidney disease

View session detail

Author Block: **A. Dawoud**[1], W. Tapper[1], N. Cross[2,1]; [1]Univ. of Southampton, Southampton, United Kingdom, [2]Salisbury NHS Fndn. Trust, Salisbury, United Kingdom
Disclosure Block: **A. Dawoud:** None.
Background: Age-related clonal haematopoiesis (CH) is a risk factor for atherosclerotic cardiovascular disease (CVD) and all-cause mortality. Chronic kidney disease (CKD) is a worldwide health problem associated with an elevated risk of CVD, but the relationship between CH and CKD is unknown. We sought to assess (i) the relationship between CH, defined by mosaic chromosomal alterations (mCA) or somatic driver mutations, and Glomerular Filtration Rate scores estimated from cystatin-C (eGFR.cys), or creatinine (eGFR.creat) used in managing CKD and (ii) the effect of CH on developing adverse outcomes defined by myocardial infarction, stroke and mortality in CKD. Methods: We focused on subjects with both SNP array, and WES in the absence of End Stage kidney Disease (ESKD) in UK Biobank (n=190,487; median age = 58y). Array allelic intensities were used to infer mCA, which were categorized as myeloid, lymphoid, or other. Likely somatic driver mutations were extracted from WES variant calls based on recurrency in Catalogue of Somatic Mutations in Cancer, pathogenicity score, and a binomial test against mean VAF. Linear and logistic regression models were used with eGFR scores as continuous trait and CKD as binary trait (eGFR 2). A Cox proportional hazard model was used to assess the risk of developing adverse outcomes. Results: CH was identified in 5,449 (2.9%) eligible subjects. CH was negatively associated with eGFR.cys ($\hat{I}^2$=-0.82, P=6.05x10^{-5}), but not other eGFR scores, and was specifically associated with CKD (n=5,449/185,038, OR=1.02, P=1.2x10^{-8}). In participants without prevalent myeloid neoplasms, eGFR.cys was associated with myeloid-related mCA (n=148, $\hat{I}^2$=-3.37, P=6.75x10^{-3}), driver mutations in myeloid neoplasia-related genes (n=3241, $\hat{I}^2$=-1.11, P=2.78x10^{-5}), and specifically *CBL, TET2, JAK2, PPM1D* and *GNB1* but not *DNMT3A* or *ASXL1*. Excluding any prior history of CVD, we confirmed the risk of adverse outcomes in myeloid CH (HR=1.57, n=338/3,078, P=1.1x10^{-12}) compared to myeloid CH-free participants (n=9174/176,944), and the higher risk in CKD: HR=2, n=1,186/6,991 P=7.6x10^{-78} compared to CKD free participants (n=8326/173,031). Within the CKD group, the risk was further increased for subjects with myeloid CH (HR=1.62, n=59/226, P=1.5x10^{-3}) compared to those who were myeloid-CH free (n=1,127/6,765). Summary/Conclusion: CH, and specifically myeloid CH, is associated with the decrease in kidney function measured by eGFR.cys. Myeloid CH stratifies the risk of adverse outcomes in CKD.

PrgmNr 2424 - A Haptoglobin (HP) Exon Deletion Polymorphism Associates with Alzheimerâ□□s Disease Status with Stratification of *APOEe4* Status

View session detail

Author Block: **H. Bai**[1], W. S. Bush[2]; [1]System Biology and Bioinformatics, Dept. of Nutrition, Case Western Reserve Univ., Cleveland, OH, [2]Case Western Reserve Univ., Cleveland, OH
Disclosure Block: **H. Bai:** None.

The free hemoglobin capturing protein - haptoglobin (HP) is distinguished into two alleles, *HP2* and *HP1*, which has a two-exon deletion. We have previously demonstrated their association with HIV-associated neurocognitive impairments. We further hypothesize that this structural variation is also associated with dementia related disorders such as Alzheimerâ□□s disease (AD). To investigate this, we first pre-imputed chromosome 16 for every participant from the Alzheimerâ□□s Disease Genetics Consortium (ADGC) that passed genotyping quality control to the TOPMed genotype marker set by cohort, followed by a subsequent imputation for the *HP* genotypes using a custom reference panel, again respectively within each cohort from ADGC. Finally, we evaluated the association between HP genotype and the AD case/control status for all the cohorts jointly by using logistic regression, adjusting for sex. We also performed analyses stratified by the apolipoprotein E (*APOE)* e*4* allele status to account for the strong effect of *APOE* on AD. In the *APOE e4* dominant group, *HP2* additively decreased the risk of AD [OR=0.943, CI=(0.916, 0.970), p=0.042, n=10,054] and this effect was not present in the non-*APOE e4* group (n=12,631) suggesting the effect may depend on an *APOE e4* context. This conclusion was also supported by congruent analyses examining the effect of *APOE e4* on AD stratified by *HP2* status, which shows a decrease in *APOE e4* effect as the *HP2* allele count increased. The *HP2* effect was observed to be larger within the more phenotypically aligned National Alzheimer's Coordinating Center (NACC) cohorts only [OR=0.921, CI=(0.890, 0.953), p=0.017, n=7,448] while the *HP1* showed a detrimental effect [OR=1.129, CI=(1.076, 1.184), p=0.013, n=7,448]. We also considered the onset of AD as an event and conducted a time-to-event study over age with a Cox proportional hazard regression model in the NACC cohorts. *HP2* alleles additively decreased the hazard of AD (Î² =-0.062Â±0.023, p=6.85e-3, n=7,448) and *HP1* dominantly increased the hazard (Î² =0.089Â±0.033, p=7.9e-3, n=7,448). The *HP1* dominant effect remained after adjusting for the *APOE e4* additive effect while the *HP2* additive effect became marginal. These results are somewhat consistent with our prior findings for neurocognitive impairment in people living with HIV (though there are obvious differences in the phenotype and age effects). Based on these analyses, the HP protein may be an imminently translatable therapeutic target for neurocognitive impairment.

PrgmNr 2425 - Deep learning-based feature extraction in neuroimaging genetics for Alzheimerâ□□s Disease

View session detail

Author Block: **D. Chakraborty**[1], H. Xue[2], Z. Zhuang[3], W. Pan[4], ADNI; [1]Sch. of Publ. Hlth., Univ. of Minnesota, Minneapolis, MN, [2]Univ. of Minnesota, Twin Cities, St Paul, MN, [3]Univ. of Minnesota, Minneapolis, MN, [4]Univ Minnesota, Minneapolis, MN
Disclosure Block: **D. Chakraborty:** None.
The prognosis and treatment of the patients suffering from Alzheimerâ□□s disease (AD) have been one of the most important and challenging problems over the last few decades. To better understand the mechanism of AD, it is of great interest to identify genetic variants associated with brain atrophy. Commonly in these analyses, neuroimaging features are extracted based on one of many possible brain atlases with FreeSurf and other popular softwares, which however may lose important information due to our incomplete knowledge about brain function embedded in these suboptimal atlases. To address the issue, we propose convolutional neural network (CNN) models applied to both three-dimensional whole-brain structure MRI data and divided multi-branch data structures to perform automatic feature extraction. These image-derived features are then used as endophenotypes in Genome-Wide Association Studies (GWAS) to identify associated genetic variants. When applied to the ADNI data, we identified several associated SNPs which have been previously shown to be related to several disorders such as depression, schizophrenia and dementia. For example, we found SNP rs2075650 which is located close to *ApoE4*, yet may independently influence risk of AD; SNP rs2196315 (P value=8.02 x 10^{-8}) corresponding to gene *ADCY8* which is known to be associated with Dissociative Amnesia; SNP rs2395095 from the *ADK* gene, known to be responsible for triggering cognitive impairment and seizures. Furthermore, the results include SNP rs173754 (P value=3.33x10^{-6}) which has been shown responsible for Attention Deficit Hyperactivity Disorder and rs9257694 (P value= 1.55 x 10^{-6}) associated with *OR14J1* (olfactory receptor family 14 subfamily J member 1) responsible for abnormalities and impaired functions of the olfactory system which are seen to appear earlier than other AD symptoms. Overall, our work suggests a novel idea of automated feature extraction and the subsequent GWAS study focusing on the entirety of brain structure, thus identifying the genes associated with Alzheimerâ□□s Disease.

PrgmNr 2426 - Evaluation of the role of common variant burden in the familial aggregation of epilepsy

View session detail

Author Block: **V. Smuk**[1], C. Leu[1], J. A. LÃ³pez-Rivera[1], L. Ferguson[2,3], L. Jehi[2], I. M. Najm[2], R. M. Busch[2,3], D. Lal[1,4]; [1]Genomic Med. Inst. & Epilepsy Ctr., Cleveland Clinic, Cleveland, OH, [2]Epilepsy Ctr., Neurological Inst., Cleveland Clinic, Cleveland, OH, [3]Dept. of Psychiatry & Psychology, Neurological Inst., Cleveland Clinic, Cleveland, OH, [4]Stanley Ctr. of Psychiatric Res., Broad Inst. of Harvard and MIT, Cambridge, MA

Disclosure Block: **V. Smuk:** None.

Rationale: Polygenic risk scores (PRS) have proven to help predict the genetic risk of developing a disease phenotype. We previously showed that common genetic risk associated with epilepsy is significantly enriched in patients with focal and generalized epilepsy. For several neurological disorders such as schizophrenia, migraine, and autism, it has been shown that patients with a known family history of the disease carry a significantly higher polygenic risk than those with no family history. However, the relationship between family history and the polygenic burden has not been established for epilepsy. Our aim was to quantify the genetic burden in epilepsy cohorts to assess if the genetic risk is further enriched in cases of familial epilepsy.

Methods: Using genome-wide genotyping data generated with the Illumina Global Screening Array-24 with Multi-disease drop-in (GSA-MD v1.0), we quantified the common genetic burden in individuals with FE or GE from a single clinical center, the Cleveland Clinic Epilepsy Center. Quality-filtered genotyping data of all individuals was imputed to the TOPMed Reference panel r2 using Minimac4 and reference-based phasing with Eagle-v2.4, as implemented on the TOPMed Imputation server. PRS were derived from summary statistics of the ILAE Consortium on Complex Epilepsies European ancestry GWAS for generalized and focal epilepsy.

Results: After quality control, we identified 336 individuals of European ancestry with FE or GE and a known family history of epilepsy (FE, 40 out of 289 individuals with FE and at least one first-degree relative; GE, 15 out of 47 individuals). Overall, we found that individuals with FE or GE and first-degree relatives with epilepsy had higher FE-PRS or GE-PRS, respectively than individuals without a family history. At the time of the conference, we will present the results of a validation cohort and the combined cohort of all individuals with FE or GE and a known family history of epilepsy.

Conclusions: Our findings demonstrate that common polygenic risk for epilepsy contributes to the familial aggregation of epilepsy. Thus, our results could be used to inform future genetic studies with a specific focus on familial cases.

PrgmNr 2427 - Exploring genome wide association signals in mesial temporal lobe epilepsy

View session detail

Author Block: **P. Mello MagalhÃ£es**[1], E. Bruxel[2], M. C. P. Athie[1], M. K. M. Alvim[1], R. Secolin[3], C. L. Yasuda[1], F. Cendes[1], I. Lopes-Cendes[4]; [1]UNICAMP, Campinas, Brazil, [2]Univ. of Campinas, Campinas, Brazil, [3]Univ. of Campinas - UNICAMP, Campinas, Brazil, [4]Univ Campinas - UNICAMP, Campinas, Brazil
Disclosure Block: **P. Mello MagalhÃ£es:** None.
Introduction: Mesial temporal lobe epilepsy (MTLE) is the most common form of focal epilepsy. MTLE is a genetically complex disease for which the identification of predisposing genes is still elusive. Recent GWAS point to a few genomic regions that may harbor genetic variants predisposing to MTLE. These regions include a locus on chromosome (ch) *2q24.3*, detected in multiple association studies. The ch *2q24.3* region includes the *SCN1A* gene and several other genes that encode ion channel subunits, which could be involved in the predisposition to MTLE. More recently, two additional loci have also been associated with MTLE: *3q25.31* and *6q22.31*. This project aims to look deeper into these candidate loci in a large cohort of patients with MTLE.
Materials and Methods: We analyzed SNPs located within the three candidate regions previously identified on chs *2q24.3*, *3q25.31*, and *6q22.31*. The p-values were adjusted for multiple comparisons.
Results: We studied 472 patients and 415 controls and identified a significant association only at the candidate locus on ch *2q24.3* (p=0.03301). Target re-sequencing of such locus is underway, and new data is expected to be available soon.
Conclusion: We found a significant genetic association for MTLE on ch *2q24.3* in our cohort of patients with MTLE. We also achieved a fine genetic mapping of the candidate region, which is being sequenced to identify putative causative variants. Our study contributes to the ongoing worldwide efforts to unravel the genetics of complex epilepsies.
Supported by: FAPESP, Brazil; CAPES, Brazil.

PrgmNr 2428 - Investigating the contribution of common variants to neurodevelopmental disorders: Expression-modifying variants and polygenic scores

View session detail

Author Block: **E. Wigdor**, E. Arciero, Q. Huang, K. E. Samocha, V. Chundru, M. E. Hurles, H. C. Martin; Wellcome Sanger Inst., Hinxton, United Kingdom
Disclosure Block: **E. Wigdor:** None.
Work from the Deciphering Developmental Disorders (DDD) study has demonstrated an excess of rare, inherited coding variants among â¼5,500 undiagnosed probands with unaffected parents and that such variants are over-transmitted to probands. This suggests these variants contribute to patientsâ disorders but are incompletely penetrant. As common variants explain 7.7% of the variance in risk for rare neurodevelopmental disorders (NDDs) (Niemi *et al*., 2018), we hypothesised that these common variants may modify penetrance of rare coding variants.
Castel *et al*. (2018) have shown that common cis-expression quantitative trait loci (cis-eQTLs) modify the penetrance of deleterious coding variants and contribute to disease risk in cancer and autism cohorts. We investigated whether common cis-eQTL increase the penetrance of inherited rare (MAF
Similarly, we investigated the role of polygenic scores (PGS) in the penetrance of rare coding variants in DD genes in probands and their parents. We found that PGS for cognitive performance, educational attainment and schizophrenia explain a small but significant (p -20) amount of the variance in risk for NDDs (on the liability scale): 0.49%, 0.61% and 0.15%, respectively, 0.85% when combined (N = 6,987 cases; 9,270 controls). We then tested whether these PGS are over-transmitted from unaffected parents to probands with rare coding variants in DD genes (MAF -11 for all comparisons).
Overall, in a large sample size (N = 3,172) of NDD patients, we see no evidence to date that cis-eQTLs or PGS modify penetrance of rare coding variants. Our power is limited by pooling variants across genes (many of which will not be pathogenic) and the low variance currently explained by multi-eQTL models and PGS. We will extend our analysis to ~10K NDD trios in Genomics England to increase our power to detect modifying effects.

PrgmNr 2429 - Mendelian Randomization analysis reveals a potential causal relationship for lipids in amyotrophic lateral sclerosis risk

View session detail

Author Block: **D. Baird**[1], W. van Rheenen[2], Y. Huang[1], M. Bakker[2], C-Y. Chen[1], J. H. Veldink[3], H. Runz[1], E. A. Tsai[1]; [1]Biogen, Cambridge, MA, [2]Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, [3]Univ Med Ctr Utrecht, Utrecht, Netherlands

Disclosure Block: **D. Baird:** Salary/Employment; Biogen Inc.

Deciphering the lipid fractions involved in disease risk is challenging due to the high correlation between the different lipid measures. Multivariable Mendelian randomization (MVMR), which models exposures together, can be used to help resolve the independent causal effects. Therefore, we applied MVMR to lipid and amyotrophic lateral sclerosis (ALS) GWAS data to investigate the role of lipids in ALS aetiology. For the lipid exposures, we used the summary statistics from UK BioBank GWAS conducted by Richardson et al 2020 on low-density (LDL-C) and high-density (HDL-C) lipoprotein cholesterol, triglycerides (TG), apolipoprotein A (ApoA) and apolipoprotein B (ApoB) measures. For the ALS outcome, we used the summary statistics from van Rheenen et al 2021 European GWAS (27,460 cases and 112,018 controls). To obtain the genetic instruments, we selected all genetic associations at P-8 across the lipid measures (after LD clumping). Instrument outliers were identified using radial analysis (Q-8 cut-off, there were 313 independent SNPs available for HDL-C, 145 for LDL-C, 264 for TG, 260 for ApoA and 168 for ApoB, from which 31 SNPs were removed as outliers from the HDL-C, 15 for LDL-C, 22 for TG, 24 for ApoA and 13 for ApoB MR analyses. Before outlier removal, the univariate MR (Inverse Variance Weighted method) suggested a relationship between increased ApoB (OR=1.09, P=0.014) and ALS risk, while other univariate results did not show significant effects (p-values: ApoA 0.144, HDL-C 0.045, LDL-C 0.053, TG 0.519). After outlier removal, evidence for a relationship between increased LDL-C and ALS also emerged (OR=1.10, P=0.0078). However, in the subsequent MVMR analysis, increased levels of ApoA (OR=1.21, P=0.21) rather than ApoB (OR=1.01, P=0.94) showed stronger evidence for an independent effect on ALS risk albeit with nonsignificant p-values. Our MR analysis suggests that lipids may have a causal role in ALS, but we are unable to reliably pinpoint the causative lipids (as the MVMR is underpowered therefore findings inconclusive). We also observed high heterogeneity in the instrument effects so further work will be needed to assess the impact of horizontal pleiotropy on all the MR findings.

PrgmNr 2430 - Polygenic risk score in Multiple Sclerosis: understanding the challenges of the post-GWAS era for complex diseases

View session detail

Author Block: **I. Charles**[1], S. Limou[1], O. Rousseau[1], V. Douillard[1], V. Mauduit[1], P-A. Gourraud[1], N. Vince[2]; [1]CRTI, UniversitÃ© de Nantes, Nantes Cedex 01, France, [2]UniversitÃ© de Nantes, Nantes Cedex 01, France

Disclosure Block: **I. Charles:** None.

Multiple Sclerosis (MS) is a complex disease with an undeniable genetic burden and an important environmental component. In the past 10 years, several MS GWAS identified more than 200 genomic regions, with the strongest signal consistently mapping to the MHC region and HLA-DRB1*15:01. Most of these genetic risk factors only have a slight effect on MS susceptibility (odds ratios (OR) around 1.1). The Multiple Sclerosis Genetic Burden (MSGB) is a polygenic risk score (PRS) developed to summarize patientâ□□s genetic risk. It is calculated on an additive log model using allelic OR as a weight for each SNP. We used GWAS data from the Welcome Trust Case Control Consortium (WTCCC, 18,859 controls and 11,376 cases) to implement a historical approach and better understand MS genetics, by calculating the MSGB using SNPs known in 2011a, 2011b, 2013, 2016 and 2019. As expected, the addition of newly discovered SNPs increased the score. We also observed:

- The 2019 score showed no significant differences between male and female, average of 20.43 and 20.36, respectively. This suggests that women higher MS risk could be explained by environmental factors rather than genetics.
- We do not observe correlation between the first principal component of the ancestry PCA (which reflects the latitude) and the score.
- The effect of the log(OR) weighting is limited by the large number of associated variants, high weighted / unweighted score correlation (r2=0.93). Similarly, the HLA SNPs effect on MSGB is also limited, high with / without HLA score correlation (r2=0.90).
- Case-control scores differences gradually increased: 6 SNP (2011a), Effect Size=0.39Â±0.01; 15 SNP (2016), ES=0.45Â±0.01; 51 SNP (2011b), ES=0.28Â±0.01; 90 SNP (2013), ES=0.34Â±0.01; 184 SNP (2019), ES=0.78Â±0.01. All p-values were below the smallest representable number -308.
- The 2019 score showed larger standard deviation between extreme individuals compared to other scores: 2011a SD=0.68, 2011b SD=0.55, 2013 SD=0.67, 2016 SD=0.75, 2019 SD=0.97.
- The MS individual with the highest 2019 score is not the same MS individual with the highest score in other years. We observe a high variation between the MSGB score of a given individual throughout the years.

PRS calculation is by definition dependent of the genetic data. Here we showed that even though MSGB is consistently higher in MS compared to controls in population, it is highly variable at the individual level. Moreover, we could not identify any genetic impact on known confounding factor in MS (sex and latitude). Overall, our results illustrate the necessity to pursue our effort in exploring PRS in MS disease to better understand genetic component of this complex pathology.

PrgmNr 2431 - Whole genome sequencing association analysis of general cognitive function in a multi-ethnic sample from the Trans-Omics for Precision Medicine (TOPMed) Program

View session detail

Author Block: **J. A. Smith**[1], M. Kho[1], J. Bressler[2], C. Sarnowski[3], W. Zhao[1], Y. Wang[1], F. Ammous[1], J. C. Bis[4], P. Nyquist[5], S. Heckbert[6], C. L. Satizabal[7], Q. Yang[8], B. M. Snively[9], A. Rodrigue[10], E. LITKOWSKI[11], D. C. Glahn[12], K. M. Hayden[13], A. Fitzpatrick[4], B. Psaty[4], S. L. Kardia[14], M. Fornage[15], S. Seshadri[16], TOPMed Neurocognitive Working Group; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ Texas Sch Pub Hlth, Houston, TX, [3]Univ. of Texas Hlth.Sci. Ctr., Houston, TX, [4]Univ. of Washington, Seattle, WA, [5]Johns Hopkins, Baltimore, MD, [6]Univ. of Washington, Ann Arbor, MI, [7]UT Hlth.San Antonio, San Antonio, TX, [8]Boston Univ Sch Publ. Hlth., Boston, MA, [9]Wake Forest Univ. Sch. of Med., Winston-Salem, NC, [10]Boston Children's Hosp., Harvard Univ., Boston, MA, [11]Univ. of Colorado, Aurora, CO, [12]IOL & Yale, Hartford, CT, [13]Duke Univ, Durham, NC, [14]Univ Michigan, Ann Arbor, MI, [15]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [16]Berlin, MA

Disclosure Block: **J.A. Smith:** None.

Background: General cognitive function is an index of performance in multiple cognitive domains. GWAS of general cognitive function, conducted mostly in European ancestry (EA) populations, have identified hundreds of genetic variants. However, investigations of rare (MAFMethods: We performed whole-genome sequence (WGS) association analysis for general cognitive function in N=12,615 EA, N=4,419 African ancestry (AA), and N=1,140 Hispanic/Latino ethnicity (HIS) participants from the NHLBI TOPMed Program, after exclusion for dementia, clinical stroke, or age =10), and SKAT-O was used for rare variant aggregation tests (MAF5 alternate alleles). Using SKAT-O, we tested coding variants within genes (high-confidence LoF, predicted deleterious missense and protein-altering) and both coding and non-coding variants within genes and their regulatory elements (enhancers and promoters). Associations were evaluated within and across ancestry/ethnicity with adjustment for age, sex, study, and genetic principal components, before and after adjustment for educational attainment. **Results**: Single-variant association analysis identified three genome-wide significant loci for general cognitive function: intronic in *AMPH* (rs17500486, MAF=3.2%, P=3.4x10^{-8}) in EA, proximal to *FTLP5* and *ATCAY* (rs570203641, MAF=0.1%, P=2.7x10^{-8}) in EA, and upstream of *RN7SL300P* (rs113037723, MAF=8.0%, P=8.8x10^{-9}) in AA. After adjustment for educational attainment, an additional two loci were identified: intronic in *SYNPO2* (rs146805942, MAF=2.1%, P=3.4x10^{-8}) in AA and intronic in *BRINP3* in HIS (rs72729138, MAF=6.6%, P=3.0x10^{-8}). Rare variant aggregation tests identified two suggestive associations using coding variants (*SH3GL3* in EA (P=3.2x10^{-6}) and *P2RY11* in AA (P=9.0x10^{-6})) and one association using coding and noncoding variants (*AC113410.3* in EA, P=9.06x10^{-6}). Of the nine genes identified, all except *RN7SL300P* and *AC113410.3* had previous evidence of association with neuronal development or activity, neurological/cognitive disorders, or cognitive function. No variants or genes were associated with general cognitive function at the genome-wide threshold in the pooled analysis of all ancestry/ethnic groups. **Conclusions**: In this WGS association analysis, we identified several new variants associated with general cognitive function, many of which lie in genes that play a role in brain function. Identified loci tended to be specific to ancestry/ethnicity.

PrgmNr 2432 - An IQ-matched genetic comparison between cases with ASD and typically-developing controls

View session detail

Author Block: **Z. Schmilovich**[1], G. HUGUET[2], E. Douard[3], C. Liao[1], J. Ross[1], G. Schumann[4], P. A. Dion[1], S. Jacquemont[5], G. A. Rouleau[6]; [1]McGill Univ., Montreal, QC, Canada, [2]CHU Ste justine, Montreal, QC, Canada, [3]UniversitÃ© de Montreal, Montreal, QC, Canada, [4]King's Coll. London, London, United Kingdom, [5]CHU Sainte justine Res. Ctr., Montreal, QC, Canada, [6]Montreal Neurological Inst.-Hosp., Montreal, QC, Canada

Disclosure Block: **Z. Schmilovich:** None.

Autism spectrum disorder (ASD) is a clinically and etiologically heterogeneous disorder that affects ~1% of the global population. The estimated heritability of ASD is high (0.65-0.91) and both rare and common variants contribute to ASD-risk. Rare de novo and inherited CNVs that substantially increase ASD-risk are present in 8-14% of cases with ASD. In fact, we have estimated that any CNV genome-wide encompassing genes intolerant to haploinsufficiency increase ASD-risk even after adjusting for their negative effects on cognition. Conversely, there is a strong (rG = 0.199) genetic correlation between ASD and IQ, such that common variants which increase ASD-risk also increase IQ in the general population.

The relationship between genetic risk, cognitive ability, and ASD is paradoxical and remains contentious. We aim to study how rare and common variants interact to confer ASD-risk, while carefully controlling for their effects on cognition using IQ-matched cases and controls.

Standard genotyping QC was performed on cases with ASD from the SPARK cohort (n = 1988) and typically-developing (TD) controls from the IMAGEN cohort (n = 1,503). Samples of inferred European ancestry were selected using the KING software and MDS parameters. Following a 1:1 matching based on NVIQ, 1251 cases and 1251 controls (mean NVIQ = 107) were included. PRS for intelligence using the Savage et al. (2018) GWAS summary statistics were computed using PRS-CS and the PLINK â□□scoreâ□□ parameter. ASD diagnosis was modelled as a function of intelligence PRS, including NVIQ, age, and sex as covariates in the logistic regression.

Our findings showed that cases with ASD had a significantly greater burden (p=1.66e-9; OR=1.32[1.21-1.45]) of polygenic risk for intelligence compared to their TD counterparts adjusted or even matched for IQ. We hypothesize that rare genetic variants (ie.: CNVs) interact with common variation increasing IQ to confer risk for ASD. We will further investigate the interaction or additive effects of these two categories of variants.

Evaluating the combined contribution of rare and common variants across IQ-matched case-control groups may reveal ASD-specific biomarkers - a long-standing aim in the field.

PrgmNr 2433 - Analysis of age- and sex-specific comorbidity patterns among brain disorders

View session detail

Author Block: **J. A. Lopez Rivera**[1,2,3], C. Leu[2,3], I. Najm[3], D. Lal[2,3,4,5]; [1]1Dept. of Molecular Med., Cleveland Clinic Lerner Coll. of Med., Case Western Reserve Univ., Cleveland, OH, [2]Genomic Med. Inst., Lerner Res. Inst., Cleveland Clinic, Cleveland, OH, [3]Epilepsy Ctr., Neurological Inst., Cleveland Clinic, Cleveland, OH, [4]Cologne Ctr. for Genomics (CCG), Med. Faculty of the Univ. of Cologne, Univ. Hosp. of Cologne, Cologne, Germany, [5]Stanley Ctr. of Psychiatric Res., Broad Inst. of Harvard and MIT, Cambridge, MA

Disclosure Block: **J.A. Lopez Rivera:** None.

Many research studies describe comorbid neurological and psychiatric disorders in individuals with a brain disorder. Shared genetic risk factors can partially explain the observed phenotypic overlap, with a higher genetic correlation between psychiatric compared to neurological disorders. However, no studies comprehensively examined the comorbidity patterns of neurological and psychiatric disorders across the common types of brain disorders. Additionally, although many studies consider age, only a few studies have examined the sex-specific comorbidity of brain disorders. To fill this knowledge gap, we assessed the co-occurrence of brain disorders across age and sex in a large healthcare system.Based on ICD-10 diagnosis codes, we identified over one million individuals diagnosed with at least one out of 20 common brain disorder types present in the aggregated health care data of eight million patients from the Cleveland Clinic healthcare system. The identified individuals were stratified into six distinct demographic groups based on age (children, adult, and elderly) and sex (male and female). For each of the 20 brain disorders and six demographic groups, we tested the enrichment of each comorbid neurological or psychiatric disorder compared to age- and sex-matched controls without brain disorders. Overall, we performed 2,280 tests and identified 327 significantly associated brain disorder pairs with an odds ratio >4. On average, each brain disorder had five significant comorbidities across all demographic groups. We replicated well-established disease-disease associations such as the co-occurrence of stroke and focal epilepsy. In line with previous evidence, we found that psychiatric disorders had, on average, a greater number of significant associations with other brain disorders compared to neurological disorders. Specifically, individuals with psychiatric disorders were enriched for other psychiatric disorders across all three age groups. We further identified distinct patterns of comorbid brain disorders for each age and sex group. For example, while generalized epilepsy was associated with autism spectrum disorder in children and adults, it was associated with Alzheimerâ□□s disease and stroke in the elderly. Finally, we observed sex-specific differential patterns of enrichment across all tested brain disorders.The goal of this research was to provide a unified resource for the age- and sex-specific clinical relationships between brain disorders. The results of our study will inform the clinical care of patients diagnosed with these conditions.

PrgmNr 2434 - Exome reanalysis of 242 Brazilian trios reveals *NPAS3* as a new possible novel autism risk locus

View session detail

Author Block: **G. Campos**[1,2], J. Y. T. Wang[2], C. I. Samogy-Costa[1,2], A. S. Girardi[2], C. GalvÃ£o[1,2], M. Zarrei[3], S. W. Scherer[3,4], M. Passos-Bueno[1,2]; [1]Dept. of Genetics and Evolutionary Biology, Univ. of SÃ£o Paulo, SÃ£o Paulo, Brazil, [2]Human Genome and Stem Cell Res. Ctr., Univ. of SÃ£o Paulo, SÃ£o Paulo, Brazil, [3]The Ctr. for Applied Genomics, The Hosp. for Sick Children, Toronto, ON, Canada, [4]Dept. of Molecular Genetics and the McLaughlin Ctr., Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **G. Campos:** None.
Autism spectrum disorder (ASD) is a complex and heterogeneous neurodevelopmental disorder (NDD) characterized by impairment in social and communication skills, in addition to restricted and repetitive behavior. In the present study, we present a reanalysis of exome sequencing data from 242 Brazilian trios attended at CEGH-CEL, in SÃ£o Paulo, Brazil, and sequenced by the Autism Sequencing Consortium. Our cohort has been evaluated by neurologists, psychiatrists, or neuropediatrician and diagnosed with ASD. We primarily address rare *de novo* variants in autism risk genes (compiled in SFARI Human Gene Module) and describe 20 *de novo* pathogenic variants, totaling an 8,26% (20/242) molecular diagnosis rate. Most of these variants take place in well-established genes, but we report a new variant (NM_003957:c.956dupT, in addition to c.1617A>C) in the newly associated gene *BRSK2* (#MIM 609236). *BRSK2* was linked to NDDs, including ASD, and to date, only nine patients were reported in the literature. We are also looking for new candidate genes associated with ASD. Currently, we are analyzing *NPAS3* (#MIM 609430), a gene that encodes a transcription factor involved in the regulation of other genes during neurogenesis; it was already associated with schizophrenia and cognitive disability, and mice lacking both *Npas3* and *Npas1* have behavioral and neurochemical alterations. *NPAS3* is absent in SFARI and is not associated with any condition in OMIM. Our patient had regression at age 1 and at the time she was evaluated had mild ASD features and motor difficulty. She harbors a *de novo* stop-gain variant (NM_001164749:c.400C>T; p.Arg134Ter) that is absent in the control population, predicted to be deleterious (CADD score >42) and results in loss of the two PAS protein domains. *NPAS3* is intolerant to null variation (pLI=1; o/e= 0.2); there is only one equivalent loss of function variant in the gnomAD database, specifically in the neuro subset. We also identified an ASD patient in the MSSNG database with a predicted splicing disruptive *de novo* variant. Therefore, we elevate *NPAS3* as a possible novel autism risk gene candidate. Identification of additional cases will be of utmost importance to confirm this hypothesis. Altogether, this work contributes to the estimation of a molecular diagnostic rate in a diverse cohort of ASD and description of new variants, since 11/20 variants were not reported before, adds clinical and molecular information to the delineation of *BRSK2* associated Mendelian condition and suggests a novel ASD candidate. Financial support: FAPESP/CEPID 2013/08028-1.

PrgmNr 2435 - Family-based whole-genome sequencing studies of psychiatric disorders in youth

View session detail

Author Block: **M. Latsko**[1], H. M. Brickman[2], S. Choi[1], E. Vargas[1], P. Brennan[1], E. Butter[3,4], D. A. Axelson[2,5,6], J. A. Bridge[4,5,7], D. C. Koboldt[1,4], R. K. Wilson[1,4], M. A. Fristad[2,5]; [1]Nationwide Childrenâ□□s Hosp. Inst. for Genomic Med., Columbus, OH, [2]Nationwide Children's Hosp. Big Lots Behavioral Hlth.Services, Columbus, OH, [3]Nationwide Children's Hosp. Child Dev. Ctr., Westerville, OH, [4]The Ohio State Univ. Dept. of Pediatrics, Columbus, OH, [5]The Ohio State Univ. Dept. of Psychiatry and Behavioral Hlth., Columbus, OH, [6]Nationwide Childrenâ□□s Hosp. Dept. of Psychiatry and Behavioral Hlth., Columbus, OH, [7]Nationwide Childrenâ□□s Hosp. Abigail Wexner Res. Inst., Columbus, OH
Disclosure Block: **M. Latsko:** None.
Severe mental health disorders are the leading cause of disability in the United States. The impact of these diseases underscores the importance of elucidating genetic markers that predispose an individual to developing a psychiatric disorder. Our work extends beyond prior research that has detected associations via GWAS or linkage analyses. We use whole genome sequencing (WGS) of family trios to identify rare variants that may underlie psychiatric disorders, including bipolar disorder, early-onset psychosis, autism, and depression/suicide. For this pilot study, we anticipate enrollment of at least 80 individuals from 20 families, half of which have already been sequenced. Eligible families are referred by physicians and behavioral health providers at Nationwide Childrenâ□□s Hospital, frequently following psychiatric inpatient admissions. Participating family members are administered structured diagnostic interviews as a measure of their own psychological well-being, as well as a comprehensive Family History Screen (FHS) to assess for psychiatric illness in the extended family. In a 17-year-old male who presented with severe early-onset bipolar 1 disorder with psychotic features, WGS revealed a de novo missense mutation (c.3501T>A) in CACNA1D, an L-type calcium channel gene that has been implicated previously in bipolar disorder. In another family, a 16-year-old female with a diagnosis of Prader-Willi syndrome (PWS) presented with severe psychosis and a maternal family history of bipolar disorder. Genome sequencing uncovered a unique pattern of mixed, maternally-inherited uniparental disomy (UPD) affecting two distinct locations along chromosome 15: the PWS critical region, and a distal region (15q25-26) implicated in bipolar susceptibility. These findings illustrate the power of whole-genome sequencing to identify rare large-effect variants contributing to severe, early-onset presentations of otherwise complex psychiatric conditions.

PrgmNr 2436 - Genome-wide rare variant score associates with morphological subtypes of Autism Spectrum Disorder

View session detail

Author Block: **A. Chan**[1], W. Engchuan[1], M. S. Reuter[2], Z. Wang[1], B. Thiruvahindrapuram[1], B. Trost[1], T. Nalpathamkalam[1], C. Negrijn[3], S. Lamoureux[1], G. Pellecchia[1], R. Patel[1], W. W. L. Sung[1], J. R. MacDonald[4], J. Howe[1], J. Vortsman[1], N. Sondheimer[1], N. Takahashi[5], J. H. Miles[6], E. Anagnostou[7], K. Tammimies[8], M. Zarrei[9], D. Merico[10], D. J. Stavropoulos[11], R. K. Yuen[4], B. A. Fernandez[12], S. W. Scherer[4]; [1]Hosp. for Sick Children, Toronto, ON, Canada, [2]Toronto, ON, Canada, [3]Eastern Hlth., St. John's, NL, Canada, [4]The Hosp. for Sick Children, Toronto, ON, Canada, [5]Thompson Ctr. for Autism and Neurodevelopmental Disorders, Columbia, MO, [6]Thompson Ctr, Univ. of Missouri, Columbia, MO, [7]Holland Bloorview, Toronto, ON, Canada, [8]Karolinska Inst.t, Stockholm, Sweden, [9]The Ctr. for Applied Genomics, The Hosp. for Sick Children, Toronto, ON, Canada, [10]Deep Genomics Inc., Toronto, ON, Canada, [11]Hosp for Sick Children, Toronto, ON, Canada, [12]Children's Hosp. Los Angeles, Los Angeles, CA

Disclosure Block: **A. Chan:** None.

Defining different genetic subtypes of Autism Spectrum Disorder (ASD) can help enable the prediction of developmental outcomes. Rare variant studies often use burden analyses to compare the frequency of rare variants, equally weighted, between cases and controls or among ASD subtypes. However, effect sizes depend on the affected gene and variant type, and these variables should be considered in rare variant analyses. Here, we used whole-genome sequencing (WGS) data and detailed clinical morphology data from two independent cohorts to: 1) develop a genome-wide rare variant score (GRVS) to measure the relationship between rare variants and dysmorphology, and 2) examine the contribution of rare and common variants in morphological ASD subtypes. Based on minor physical and major congenital anomalies, we categorized 325 Canadian children into dysmorphic (n=138) and non-dysmorphic (n=181) ASD. We identified clinically significant single nucleotide variants, short insertions/deletions and copy number variants in 14.2% (46/325) of ASD cases. The yield of clinically significant variants (CSVs) was significantly higher in dysmorphic (25.9%; 35/135) compared to non-dysmorphic (5.8%, 11/190, p=8.7Ã□10^{-5}) ASD. We developed a method to calculate a GRVS from WGS in each individual. GRVS is a weighted sum of the number of variants in morphology-associated coding and non-coding regions, weighted by their effect sizes. Probands with dysmorphic ASD had a significantly higher GRVS compared to those with non-dysmorphic ASD (*P*= 0.027). Using the GRVS formula, we calculated a score for CSVs that overlapped a morphology-associated region. In 47% of probands with CSVs, CSVs contributed to the majority (>50%) of GRVS. When we excluded the 46 probands with CSVs, those with dysmorphic ASD still had significantly higher average GRVSs than those with non-dysmorphic ASD (*P*= 0.048). Using the polygenic transmission disequilibrium test, we observed an over-transmission of ASD-related common variants in non-dysmorphic ASD probands (*P*= 2.9Ã□10^{-3}). We replicated our findings using WGS data from 442 ASD probands that had accompanying morphology data from the Simons Simplex Collection. Our findings provide support for an alternative genomic classification of ASD subgroups using morphology data, which may inform intervention protocols. The GRVS method can also be applied to future studies to better measure the full impact of rare variants on susceptibility to ASD and other complex disorders.

PrgmNr 2437 - Integrating Brain Imaging Phenotypes, Genomics, and Substance Use

View session detail

Author Block: **Y. Chang**[1], J. Bijsterbosch[1], A. Anokhin[1], R. C. Culverhouse[2], L. J. Bierut[3]; [1]Washington Univ. in St.Louis, St.Louis, MO, [2]Washington Univ Med Sch, St Louis, MO, [3]Washington Univ Sch Med, St Louis, MO
Disclosure Block: **Y. Chang:** None.
Cigarette smoking is associated with persistent neurochemical and pathological changes in brain. Recently UK Biobank retrieved MRI data from healthy individuals which allows for the opportunity to identify pre-symptomatic markers for neuropathological conditions associated with smoking behaviors.
Using the currently available subset of UK Biobank data of 19,301 individuals, we first studied the association between cigarette smoking and brain imaging-derived phenotypes (IDPs). We used total of 359 IDPs (305 structural MRI-derived phenotypes and 54 diffusion MRI-derived phenotypes). Smokers were divided into 1) Current daily smokers (n=799), former daily smokers (n=4,324), former occasional smokers (n=2,385), non-smokers who smoked less than 100 cigarettes (n=3,290) and never smokers (n=8,247). We performed regression analyses to measure the association between 359 IDPs and 1) smoking status and 2) pack years. Then we created smoking-related polygenic risk scores (PRSs) using summary statistics from GWAS & Sequencing Consortium of Alcohol and Nicotine use. After assessing its predictive ability for UK Biobank smoking phenotypes, we performed regression analyses to observe the association between PRSs and IDPs.
Current daily smokers (compared to never smokers) had strong association after correcting for multiple testing with 48 of 359 IDPs, former daily smokers had 72 associations, former occasional smokers had 1 association, and non-smokers had none. The brain regions most significantly associated were caudate, putamen, amygdala, pallidum and thalamus for both current and former daily smokers. Tract posterior thalamic radiation was most significant for the two groups. The statistical significance for these regions and the tract decreased for former occasional smokers. The dose-effect of smoking (pack years) generally showed that higher dose of cigarettes was associated with decreased brain volume and white matter integrity. For PRSs and IDPs, smoking cessation PRS was significantly associated with volumes in putamen, pallidum and accumbens while cigarettes per day PRS was associated with volume in putamen and smoking initiation PRS with volume in pallidum. This study aims to establish a link between genetics, brain, and smoking behavior. By dividing the smokers into different groups and observing the dose-effect of cigarettes, we concluded that depending on the smoking dosage, the brain can either be recovered or permanently affected following smoking cessation. Also, PRS showed how genetic predisposition to different stages of smoking behavior can affect different regions of the brain.

PrgmNr 2438 - Interactions between variants in detoxification and permeability barrier genes and xenobiotics in Autism Spectrum Disorder

View session detail

Author Block: **J. Xavier Santos**[1,2], A. Marques[1,2], C. Rasga[1,2], H. Martiniano[1,2], J. Vilela[1,2], A. M. Vicente[1,2]; [1]NIH Doutor Ricardo Jorge, Lisbon, Portugal, [2]Faculdade de CiÃªncias, BioISI - Biosystems & Integrative Sci. Inst., Univ.e de Lisboa, Lisbon, Portugal
Disclosure Block: **J. Xavier Santos:** None.
Introduction: While heritability estimates of 50-80% suggest that gene-environment interactions may contribute to the etiology of Autism Spectrum Disorder (ASD), there are insufficient studies simultaneously addressing both components. Thus, we implemented the GEnvIA project, aiming at the integration of genetic variants and early-life exposure data collected from ASD-subjects. The objectives of the current work were: 1) to identify Single Nucleotide Variants (SNVs) in a panel of 18 genes (the XenoReg gene panel), that regulate detoxification and permeability barriers (placenta and blood-brain barrier - BBB) processes, in ASD subjects; 2) to search for interactions between these genes and xenobiotics implicated in ASD. **Methods:** Samples from 218 ASD subjects were sequenced using Illumina platforms. GATK best practices were applied for read alignment and variant calling. Variant functional annotation was done with Variant Effect Predictor tool. Only rare (gnomAD MAF in silico were considered. The Comparative Toxicogenomics Database (CTD) was interrogated for gene-environment interactions. **Results:** In 60/218 (27.5%) subjects, we identified 55 unique variants (4 LoF and 51 missense) in 16 out of the 18 XenoReg genes. These included *CYP2A13*, *CYP4X1*, *NQO1*, *STS*, *UGT2B10* and *UGT2B15,* which encode enzymes responsible for the metabolism of xenobiotics; *ABCA8, ABCB1, SLC3A2* and *SLC22A5,* which encode transporters controlling the flux of toxins across the BBB and placenta; *CLDN3* and *TJP3* that encode components of BBB tight junctions; and *KIF17* a kinesin likely involved in the formation of the placenta. Based on CTD query we constructed a novel gene-environment interactions network putatively relevant for ASD, revealing that all 16 genes interact with xenobiotics implicated in ASD, including benzo(a)pyrene, bisphenol A, particulate matter, phthalates, pesticides and valproic acid. These are ubiquitous neurotoxins, present in everyday products, to which exposure can, at some extent, be mitigated. **Discussion:** Given the extreme vulnerability of the developing brain to toxins, subjects with variants in XenoReg genes might be at higher risk of ASD onset due to early-life exposure to ubiquitous xenobiotics even at permitted concentrations. The dysregulation of detoxification and barrier permeability processes may lead to the buildup of toxins in the immature brain, culminating in neuropathological processes (eg. epigenetics, oxidative stress, neuroinflammation and endocrine disruption) that impair neuronal, synaptic and chromatin remodeling functions associated with ASD.

PrgmNr 2439 - Polygenic scores for cortical and subcortical brain structures are associated with cognition

View session detail

Author Block: **T. Silzer**[1], C. Burton[1], M. Lemire[1], P. Arnold[2], R. Schachar[1], J. Crosbie[1]; [1]Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of Calgary, Calgary, AB, Canada
Disclosure Block: **T. Silzer:** None.
Background. Variation in brain structures has been linked to neuropsychiatric disorders including schizophrenia, attention deficit-hyperactivity disorder (ADHD) and autism spectrum disorder. Across these disorders, cognitive deficits are common. Genetics plays an important role in determining both brai structure and cognition; however, a gap exists in our understanding of the potential shared genetic variation between brain structure and cognitive function. The purpose of the study was to test if polygenic scores (PGS) derived from GWAS of global and regional brain structures (e.g. surface area and thickness) were associated with aspects of cognition (e.g. response variability). **Methods.** PGS were computed in the population-based Spit for Science sample of children and youth (N=~5200) from summary statistics of genome-wide meta-analyses of global and regional cortical and sub-cortical surface area and thickness (N=~30,000) published by the Enhancing Neuro Imaging Genetics Through Meta-Analysis (ENIGMA) consortium. PGS were generated using PRSice v1.25. PGS were tested for association with aspects of cognition (e.g. response variability), as measured by the Stop Signal Task in Spit for Science. **Results and Conclusion.** Preliminary results indicate a lack of global brain structure correlates but a presence of regional associations with cognition. However, significant brain region associations suggest that genetic variation underlying specific cortical and sub-cortical structures may be linked to key features of neuropsychiatric disorders. Study findings will have direct implications for understanding the potential shared biology of brain structure and cognition, and the genetic etiology of neuropsychiatric disorders that feature deficits in cognition. Future works will also investigate the genetic overlap between regional brain volume and cognition.

PrgmNr 2441 - Trans-ancestry GWAS meta-analysis of alcohol and tobacco addiction in 3.4 million individuals

View session detail

Author Block: **G. Saunders**, GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN), Trans-Omics for Precision Medicine (TOPMed); Univ. of Minnesota, Minneapolis, MN
Disclosure Block: **G. Saunders:** None.
The use and abuse of nicotine and alcohol accounts for >100 million disability-adjusted life years across the globe, constituting one of the worldâ□□s leading public health problems. Despite this, the vast majority of genome-wide association studies (GWAS) thus far have been restricted to individuals of European ancestry, representing

PrgmNr 2442 - A genome-wide association study of MÃ©niÃ¨re disease in the UK Biobank cohort reveals one genome-wide significant locus and genetic correlation with mood/anxiety-related traits

View session detail

Author Block: **M. E. Bailey**[1], F. Casanova[2], L. Assad[2], D. M. Lyall[1], J. P. Pell[1], D. J. Smith[3], B. Cullen[1], G. A. J. Morrison[4], J. Ward[1], J. Tyrrell[2]; [1]Univ. of Glasgow, Glasgow, United Kingdom, [2]Univ. of Exeter, Exeter, United Kingdom, [3]Univ. of Edinburgh, Edinburgh, United Kingdom, [4]Formerly: Guy's & St. Thomas' Hosp. NHS Fndn. Trust, London, United Kingdom
Disclosure Block: **M.E. Bailey:** None.
MÃ©niÃ¨re disease (MD) has been a biological conundrum since its recognition in 1861. MD is a moderately common, complex disorder characterised by episodes of vertigo/nausea and fluctuating but progressive hearing impairment, often accompanied by tinnitus. It affects approx. 1/2,000 people of European origin (females > males) and is associated pathologically with endolymphatic hydrops. MD is often associated with considerably reduced quality of life and there are few effective non-surgical therapies. It is often confused with other vestibular and migraine-related disorders, and is comorbid with a variety of neurological and psychiatric conditions, including depression. The aetiology of MD is largely unclear. Previous HLA association reports are thus far uncorroborated in large cohorts. About 95% of cases are sporadic, with 5% familial. There are a few, thus far unreplicated, reports of rare, high-penetrance variants/loci associated with MD status within families or in sporadic cases. There are no reliable estimates of heritability. We hypothesized a polygenic component to MD in sporadic cases and carried out a case/control GWAS in UK Biobank to elucidate causal loci and aspects of the genetic epidemiology of MD.

Our analysis used 2,594 likely MD cases ascertained via either self-report, hospital inpatient records or primary care records, and approx. 441K controls excluding individuals with a history of vestibular problems. GWAS analysis using SAIGE (adjusting for age, sex, genotyping chip and evaluation centre) revealed one genome-wide significant locus in Chr. 2q14.1 (lead SNP: rs3977027; MAF = 0.15; $p = 2.9 \times 10^{-8}$), within an intron of *CBWD2*. Ancillary and downstream analyses are ongoing, but there is evidence for a small polygenic contribution, though SNP heritability estimates are low (g values 0.25 - 0.41; FDR-adj. $p = 0.020$ - 0.034), negatively genetically correlated with educational attainment (r_g -0.34; FDR-adj. $p = 0.034$) and nominally correlated with coeliac disease. These findings support a biological basis for MD, but with a modest genetic component, and establish a biological connection, based on pleiotropy or mediated effects, between MD and mood/anxiety-related brain functionality. Further understanding of the biology of MD will inform the development of more effective preventative and therapeutic approaches aimed at improving quality of life for sufferers.

PrgmNr 2443 - A phenotype-driven machine-learning model to assess the pathogenicity of digenic variant combinations: a case study on ciliopathies

View session detail

Author Block: **F. De Paoli**[1,2], I. Limongelli[1], S. Zucca[1], F. Baccalini[2], V. Serpieri[3,4], F. D'Abrusco[4], E. M. Valente[3], P. Magni[2]; [1]enGenome srl, Pavia, Italy, [2]Dept. of Electrical, Computer and BioMed. Engineering, Univ. of Pavia, Pavia, Italy, [3]IRCCS Fondazione Istituto Neurologico Nazionale Casimiro Mondino, Pavia, Italy, [4]Dept. of Molecular Med., Univ. of Pavia, Pavia, Italy
Disclosure Block: **F. De Paoli:** None.
Background:
For decades, the inheritance mechanism of genetic disorders was explained through the '*one-gene, one-disease*' paradigm according to which mutations affecting a single gene could be causative of several rare Mendelian disorders. Recently, more complex genetic models have been proposed to explain a series of genetic disorders which could not be solved through a single causative mutation, such as digenic or oligogenic inheritance. In the digenic model, two different mutated genes concur to cause the patient's phenotype. Here, we propose a method (DIVAs) to assess the pathogenicity of digenic variant combinations.
Methods:
DIVAs is a sophisticated machine learning model developed to classify combinations of variants on two different genes according to patient's phenotypes. The input to this model is the list of variants identified in the patient and a set of Human Phenotype Ontology terms describing his medical condition. Model features capture all the characteristics of a variant combination and its association with patient's phenotypes at variant, gene and gene-pair level. The output reports the list of all possible combinations, ranked according to their probability of causation. The algorithm was trained and tested on a dataset of validated pathogenic (http://dida.ibsquare.be, internal database and public literature review) and benign (https://www.internationalgenome.org/) variant combinations.
Results:
DIVAs was preliminary validated on an independent dataset of 78 cases published in the scientific literature (sensitivity 82%). The algorithm was then tested on 4 solved digenic clinical cases for which WES data were available. All the cases were related to the wide and genetically heterogeneous disease spectrum of ciliopathies whose genetic diagnosis could be driven by an oligogenic inheritance pattern. First, two male siblings with complex traits such as microcephaly, intellectual disability and seizures have been analyzed through our algorithm: DIVAs correctly classified the pathogenic variant combination involving *FRMPD4* and *PAK3* genes and ranked it in the first position. Second, DIVAs correctly prioritized the *TMEM67-MKS1* digenic combination which explains the complex traits of two fetuses affected by Meckel-Gruber syndrome, a lethal form of ciliopathy. DIVAs was also tested on unsolved cases of patients with suspected ciliopathy and results interpretation is ongoing.In conclusion, our method proved to be reliable to identify causative digenic variant combinations, contributing to uncover the missing heritability of rare genetic disorders.

PrgmNr 2444 - Associated anomalies in cases with oral clefts

View session detail

Author Block: **C. Stoll**; Laboratoire de Genetique Med.e, Strasbourg, France
Disclosure Block: **C. Stoll:** None.
Cases with oral clefts (OCs) often have other associated congenital anomalies. The reported prevalence and the types of associated anomalies vary between different studies. The purpose of this investigation was to assess the prevalence and the types of associated anomalies in a geographically well defined population. The prevalence and the types of associated anomalies in cases with OCs were collected in all live births, stillbirths and terminations of pregnancy between 1979 and 2007 in 387,067 consecutive births in the area covered by our population-based registry of congenital anomalies. Of the 789 OCs cases ascertained during this period (prevalence of 20.4 per 10,000 births), 39.5% had associated non-OCs anomalies. Associated anomalies were more frequent in cases with cleft palate (52.4%) than in cases with cleft lip and palate (37.3%) and in cases with cleft lip only (16.8%). There were 94 (11.9%) cases with chromosomal abnormalities, including 27 trisomies 13, 15 trisomies 18, 12 22 q11.2 deletion, and 40 other chromosomal abnormalities, 38 (4.8%) cases with non-chromosomal recognizable conditions including syndromes, associations, spectrums and sequence and 180 cases (22.8%) with multiple congenital anomalies (MCA). Anomalies in the musculoskeletal system (16.7%), the central nervous system (15.0%), the urogenital system (13.7%), the cardiovascular system (8.6%), and the digestive system (6.6%) were the most common MCA. The overall prevalence of associated anomalies emphasizes the need for a thorough investigation of cases with OCs. A routine screening for other congenital anomalies need to be considered in infants and in fetuses with OCs.

PrgmNr 2445 - Characteristics of keratoconus corneal epithelium in multi-omic approach

View session detail

Author Block: **K. Jaskiewicz**[1], M. Maleszka-Kurpiel[2,3], M. Rydzanicz[4], E. Matuszewska[5], J. Matysiak[5], R. Ploski[4], M. Gajecka[1,6]; [1]The Inst. of Human Genetics of the Polish Academy of Sci., Poznan, Poland, [2]Optegra Eye Hlth.Care Clinic, Poznan, Poland, [3]Chair of Ophthalmology and Optometry, Poznan Univ. of Med. Sci., Poznan, Poland, [4]Dept. of Med. Genetics, Med. Univ. of Warsaw, Warsaw, Poland, [5]Dept. of Inorganic and Analytical Chemistry, Poznan Univ. of Med. Sci., Poznan, Poland, [6]Chair and Dept. of Genetics and Pharmaceutical Microbiol., Poznan Univ. of Med. Sci., Poznan, Poland
Disclosure Block: **K. Jaskiewicz:** None.
Introduction: Keratoconus (KTCN) is a corneal disease, characterized by multifaceted etiology with genetic heterogeneity. KTCN affects 1:2000 individuals worldwide and causes substantial vision impairment on account of increased thinning. Due to pathological cone formation, the corneal epithelium (CE) in KTCN patients demonstrates a specific doughnut pattern, therefore we aimed to implement a multi-omic approach to assess in detail three distinct CE regions.
Materials and Methods: The CE samples collected during the cross-linking procedure in X KTCN patients and refractive error correction in Y myopia individuals as non-KTCN controls were separated into central, middle, and periphery regions regarding cone location. All individuals underwent detailed ophthalmological examination. The RNA/DNA/protein samples were extracted simultaneously (RNA/DNA/Protein Purification Plus MicroKit, Norgen Biotek) towards WGS (30x coverage), RNAseq (TruSeq Stranded Total RNA LibraryPrep Gold, 100mln reads/sample), and MALDI-TOF MS analyses. The obtained data were integrated.
Results: The average CE thickness values of the three analyzed regions were found to be statistically different in KTCN patients showing a characteristic doughnut pattern. The GO analysis of the 3 youngest patientsâ gene variants points to changes in pathways of immune responses and extracellular matrix composition. The effect of sex on transcriptomic and proteomic profiles was revealed. Still, we identified differentially expressed pathways involved in an extracellular matrix organization, collagen-containing extracellular matrix, and apoptosis. Characteristics of the KTCN proteomic profile were obtained.
Conclusions: The results allowed to characterize the KTCN cone corneal regions, and further pointed to the KTCN genetic heterogeneity. The differences found between the three CE regions reflect the thickness abnormalities in the KTCN cone.
Support: The National Science Centre grant no.2018/31/B/NZ5/03280.

PrgmNr 2447 - Extreme Expression Genotypes revealed regulatory effects on gene expression linked to traits and footprints of positive selection

View session detail

Author Block: **A. Sartori**[1], T. Dupuis[2], A. Brown[3], E. T. Dermitzakis[4]; [1]Genetic Med. and Dev., Univ. of Geneva, Geneva, Switzerland, [2]Population Hlth.and Genomics, Univ. of Dundee, Sch. of Med., Dundee, United Kingdom, [3]Univ. of Dundee, Dundee, United Kingdom, [4]Univ. of Geneva, Geneva, Switzerland
Disclosure Block: **A. Sartori:** None.
Rare variants can have effects on gene expression with mendelian-type consequences for disease risk. Similarly, we have now observed that genes can be regulated by multiple independent common variants (independent *cis*-eQTLs), with 686 genes identified in the latest GTEx release as regulated by 4 or more *cis*-eQTLs. In this study, we hypothesize that combinations of *cis*-eQTL alleles with common direction of effect could show as extreme cumulative effects as those observed for rare variants, and with similar downstream consequences on disease risk. We refer these multiple consistent eQTL alleles as Extreme Expression Genotypes (EEG). Using eQTLs from 49 GTEx tissues, we defined EEGs for genes with 2 or more eQTLs and tested them for association with BMI on 405719 individuals in the UK Biobank cohort. BMI was chosen as a robust polygenic phenotype, with an important causal role in disease. We found 61 significant associations between predicted extreme expression and BMI, 45 extreme predicted high expression and 16 with extreme predicted low gene expression. Among the top BMI-associated genes for high EEG there were MON1A (p = 3.04E-11), GDPD3 (p = 7.92E-10) and TMEM19 (p = 1.52E-09). The top associated low EEG effect genes were RBM6 (p =5.8E-11), GBA (p = 1.16E-08) and EIF3CL (p = 2.90E-08). We are currently investigating for footprints of positive selection on these EEHs by applying Extended Haplotype Homozygosity based - statistics.

PrgmNr 2448 - From GWAS to GWANN: Genome Wide Artificial Neural Networks and the interpretability of non-coding associations

View session detail

Author Block: **R. van der Winden**[1], A. van Hilten[1], A. Lal[2], L. Slim[2], J. Israeli[2], W. J. Niessen[1,3], H. W. Clifford[2], G. V. Roshchupkin[1]; [1]Erasmus MC, Rotterdam, Netherlands, [2]NVIDIA, Santa Clara, CA, [3]Delft Univ. of Technology, Delft, Netherlands
Disclosure Block: **R. van der Winden:** None.
While we have learned that many human traits and health conditions have genetic components, our understanding of the underlying processes is still limited due to the complexity of polygenic phenotypes. Approaches like GWAS have shown success in identifying genetic variants associated with various traits, however, most of those variants are non-coding and, in some way, regulate gene expressions. Non-coding genome annotation resources can give more knowledge about disease etiology and individual differences in humans, but we are still far away from understanding the complex relationships between genetic variants and traits.
More advanced modeling methods, such as artificial neural networks, are beginning to emerge with the aim of tackling such complexity. Several of these, such as our GenNet framework and Microsoftâ□□s Biologically Annotated Neural Networks, aim to balance complexity with interpretability by using architectures that are based on biological annotations. This somewhat solves the issues of understanding the underlying biology, but these models are not yet truly genome-wide, being limited to coding regions due to computational burdens.
Here, we present an expanded GenNet framework which links non-coding variants to nearby gene models. The network uses the whole genotype array for each individual as input (~780,000 variants). Subsequently, each variant is connected to their nearest gene to construct a second network layer of 23,654 neurons, which represent genes. We have trained the network on several traits from the UK Biobank cohort, such as hair color (9,615 cases and 9,718 controls) and type 2 diabetes (12,619 cases and 12,812 controls).
Training the network on a single GPU (NVIDIA Tesla K40m) took an average of 2.8 hours. For red hair, the model reaches an area under the curve (AUC) of 0.95 and an accuracy of ~87%. For diabetes type 2, the AUC is 0.62 with an accuracy of ~59%, versus an AUC of 0.54 achieved with the exome-based network. The five most important genes for the identification of diabetes type 2 are *GRM7*, *BRCA2*, *NYAP2*, *RBFOX1* and *LDLR*. Four out of five of these genes have previously been linked to type 2 diabetes.
Here we have demonstrated a proof of concept of the extended GenNet framework, which uses all variants on a genotype array and shows greater predictive power compared to other methods.

PrgmNr 2449 - Fusion of multiplex whole-exome and low-depth whole-genome sequencing for cost-effective rare variant assessment in the genome-wide association studies

View session detail

Author Block: **D. Taliun**[1,2], C. BhÃ©rer[1,2], R. Eveleigh[3,4], P. Nadukkalam Ravindran[3,2], C. Wei[1,2], J. St-Cyr[3], A. Paccard[3], N. Bader Asbah[3], P. Lepage[3], M. Bourgey[3,4], G. Bourque[3,4,1,5], J. Ragoussis[3,1], v. Mooser[3,2,1]; [1]McGill Univ., Montreal, QC, Canada, [2]Canada Excellence Res. Chair (CERC) Program in Genomic Med. at McGill Univ., Montreal, QC, Canada, [3]McGill Genome Ctr., Montreal, QC, Canada, [4]Canadian Ctr. for Computational Genomics (C3G), Montreal, QC, Canada, [5]Kyoto Univ., Kyoto, Japan
Disclosure Block: **D. Taliun:** None.
Rare genetic variants (minor allele frequency We sequenced an Ashkenazi Jewish trio (HG002, HG003, and HG004) using: (i) traditional WES at ~96X and ~84X in coding and UTR regions, respectively; (ii) $WEGS_{2X}$ â 4-plex WES at ~110X in coding regions and WGS at ~2X; (iii) $WEGS_{5X}$ â 4-plex WES at ~110X in coding regions and WGS at ~5X. We increased the coding region coverage in WEGS to compensate for the increase in GC bias due to multiplexing. We performed SNV and indel calling following the GATK4 best practices pipeline and evaluated precision and recall against high confidence variant calls from the Genome in a Bottle Consortium. The SNV calling precision in coding regions was concordant between WEGS (~0.996) and WES (~0.997), while recall was slightly lower in $WEGS_{2X}$ and $WEGS_{5X}$ (~0.983 and ~0.985 vs. 0.995, respectively). WES had higher rates of both precision (~0.957) and recall (~0.980) for indels in coding regions compared to $WEGS_{2X}$ (~0.918 and ~0.960, respectively) and $WEGS_{5X}$ (~0.920 and ~0.965, respectively). Although $WEGS_{5X}$ showed a slight increase in variant calling accuracy in coding regions compared to $WEGS_{2X}$, it detected more SNVs (~2,1M vs. ~1,5M) and more indels (339K vs. 215K) in non-coding regions and showed better results in downstream genotype imputation (~3,6M vs ~3,1M imputed SNVs with imputation quality r2>0.3).

In summary, we experimentally demonstrate that WEGS maintains similar precision and recall in the discovery of short coding variants compared to traditional WES, but at the same time allows for the assessment of low-frequency and common variants in non-coding part of the genome. Because WEGS pools 4 samples simultaneously and uses smaller UTR-free target sequence capture, this approach could serve as a cost-effective alternative between traditional WES and high-coverage WGS in genetic association studies.

PrgmNr 2450 - Genetically independent phenotype analysis identifies LPA and VCAM1 as drug targets for human ageing

View session detail

Author Block: **P. R. Timmers**[1,2], E. S. Tiys[3,4], S. Sakaue[5,6,7,8], M. Akiyama[9,10], T. T. Kiiskinen[8,11], W. Zhou[12,8,13], S-J. Hwang[14,15], C. Yao[14,15], The Biobank Japan Project, FinnGen Study, J. Deelen[16,17], D. Levy[14,15], A. Ganna[8,11,12], Y. Kamatani[9,18], Y. Okada[6], P. K. Joshi[2], J. F. Wilson[1,2], Y. A. Tsepilov[3,19]; [1]MRC Human Genetics Unit, MRC Inst. of Genetics & Molecular Med., Univ. of Edinburgh, Edinburgh, United Kingdom, [2]Ctr. for Global Hlth.Res., Usher Inst., Univ. of Edinburgh, Edinburgh, United Kingdom, [3]Lab. of Theoretical and Applied Functional Genomics, Novosibirsk State Univ., Novosibirsk, Russian Federation, [4]Lab. of Glycogenomics, Inst. of Cytology and Genetics, Novosibirsk, Russian Federation, [5]Ctr. for Data Sci., Harvard Med. Sch., Boston, MA, [6]Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., Suita, Japan, [7]Div.s of Genetics and Rheumatology, Dept. of Med., Brigham and Womenâ□□s Hosp., Harvard Med. Sch., Boston, MA, [8]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, [9]Lab. for Statistical and Translational Genetics, RIKEN Ctr. for Integrative Med. Sci., Yokohama, Japan, [10]Dept. of Ophthalmology, Graduate Sch. of Med. Sci., Kyushu Univ., Fukuoka, Japan, [11]Inst. for Molecular Med. Finland, Univ. of Helsinki, Helsinki, Finland, [12]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [13]Stanley Ctr. for Psychiatric Res., Broad Inst. of Harvard and MIT, Cambridge, MA, [14]Framingham Heart Study, Framingham, MA, [15]Population Sci. Branch, Div. of Intramural Res., Natl. Heart, Lung, and Blood Inst., NIH, Bethesda, MD, [16]Max Planck Inst. for Biology of Ageing, Cologne, Germany, [17]Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), Univ. of Cologne, Cologne, Germany, [18]Lab. of Complex Trait Genomics, Dept. of Computational Biology and Med. Sci., Graduate Sch. of Frontier Sci., The Univ. of Tokyo, Tokyo, Japan, [19]Lab. of Recombination and Segregation Analysis, Inst. of Cytology and Genetics, Novosibirsk, Russian Federation

Disclosure Block: **P.R. Timmers:** None.

The length and quality of life is important to us all, yet identification of promising drug targets for human ageing using genetics has had limited success. Here, we combine six large European-ancestry genome-wide association studies (GWAS) of human ageing traitsâ□□healthspan, father and mother lifespan, exceptional longevity, frailty index, and self-rated healthâ□□in a principal component framework which maximises their shared genetic architecture. The first principal component (GIP1) is more heritable than the original studies and shows strong genetic correlations with length of life as well as multiple indices of mental and physical wellbeing. We identify 27 genomic regions associated with GIP1, and provide additional, independent evidence for an effect on human ageing for loci near *HTT* and *MAML3* using a study of Finnish and Japanese subject survival. Across the genome, GIP1 associations are enriched in genes involved in haem metabolism and pathways related to transcription, neurogenesis, homeostasis, proteolysis, intracellular signalling, immunity, and the muscle system. Finally, using proteome-wide two-sample Mendelian randomisation and colocalisation, we provide robust evidence for a detrimental effect of blood levels of apolipoprotein(a) (LPA) and vascular cell adhesion molecule 1 (VCAM1) on GIP1. Together, our results demonstrate that combining multiple ageing traits using genetic principal components enhances power to detect biological targets for human ageing.

PrgmNr 2451 - Genome-wide investigation of the interplay between age-related hearing loss, tobacco smoking, and noise

View session detail

Author Block: **F. DE ANGELIS**[1,2], O. Zeleznik[3], F. Wendt[4,1,2], G. Pathak[5,1,2], A. De Lillo[1,2], D. Koller[4,6,2], B. Cabrera Mendoza[7,6,2], G. Curhan[8,9], S. Curhan[3,9], R. Polimanti[10,6,2]; [1]Dept. of Psychiatry, Yale Univ. Sch. of Med., West Haven, CT, New Haven, CT, [2]Veteran Affairs Connecticut Hlth.care System, West Haven, CT, West Haven, CT, [3]Channing Div. of Network Med., Dept. of Med., Brigham and Women's Hosp., Boston, Massachusetts, USA., Boston, MA, [4]Yale Sch. of Med., West Haven, CT, [5]VA CT Hlth.care system, West Haven, CT, [6]Dept. of Psychiatry, Yale Univ. Sch. of Med., West Haven, CT, New haven, CT, [7]West Haven, [8]Channing Div. of Network Med., Dept. of Med., Brigham and Women's Hosp., Boston, Massachusetts, USA, Boston, MA, [9]Harvard Med. Sch., Boston, Massachusetts, USA., Boston, MA, [10]New Haven, CT

Disclosure Block: **F. De angelis:** None.

In the US, approximately 33% of the population aged 65 to 74 suffers from hearing loss. Although age-related hearing loss has been linked to several adverse physical and mental health outcomes, information on genetic susceptibility and how it moderates the effect of known environmental risk factors is limited. We conducted a large-scale genome-wide association study of self-reported hearing difficulties among >400,000 unrelated participants of European descent in the UK Biobank (ages 37-73). Eighty-three independent genome-wide significant (GWS) loci (PKLHDC7B rs36062310 (P=2.63E-32). Many of GWS variants mapped to genes regulating the development and function of the inner ear. Some of the loci identified have also been previously associated with personality disorders and psychological traits (e.g., rs483143 with anxiety, neuroticism, and unipolar depression; rs12938775 neuroticism and depressed affect). The gene ontology for the loci associated with self-reported hearing difficulty were enriched in processes involved in sensory perception (GO:0050954, P=1.14E-05), nervous system synaptic processes (GO:1905244, P=7.26E-06), and cytoskeletal morphology (GO:0014731, P=2.81E-05). To further investigate hearing loss biology, we conducted a multivariate gene-by-environment analysis, testing simultaneously the interactive effects of traits related to noise and tobacco smoking. Among the GWS loci, several variants showed interactive effects with the noise-smoking covariance matrix (e.g., rs2173109, rs55938136, and rs1170465). In summary, our findings provide the first genome-wide evidence of the complex interplay linking genetic variation and environmental risk factors in the context of age-related hearing loss.

PrgmNr 2452 - Global Biobank Meta-Analysis Initiative: A genome-wide association meta-analysis identifies novel primary open-angle glaucoma loci and shared biology with vascular mechanisms

View session detail

Author Block: **V. Lo Faro**[1], J. Hirbo[2,3], A. Bhattacharya[4], W. Zhou[5,6,7], Global Biobank Meta-analysis Initiative, N. Jansonius[8], H. Snieder[9], K. Joos[10], N. Cox[2,3]; [1]Dept. of Ophthalmology, Univ. of Groningen, Univ. Med. Ctr. Groningen, Groningen, Netherlands, [2]Div. of Genetic Med., Vanderbilt Univ. Sch. of Med., Nashville, TN, [3]Vanderbilt Genetics Inst., Nashville, TN, [4]Dept. of Pathology and Lab. Med., David Geffen Sch. of Med., Univ. of California, Los Angeles, CA, [5]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [6]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, [7]Stanley Ctr. for Psychiatric Res., Broad Inst. of Harvard and MIT, Cambridge, MA, [8]Dept. of Ophthalmology, Univ. Med. Ctr. Groningen, Groningen, Netherlands, [9]Dept. of Epidemiology, Univ. of Groningen, Univ. Med. Ctr. Groningen, Groningen, Netherlands, [10]Vanderbilt Eye Inst., Vanderbilt Univ. Sch. of Med., Nashville, TN

Disclosure Block: **V. Lo Faro:** None.

Primary open-angle glaucoma (POAG) is a complex eye disease characterized by progressive loss of optic nerve function that if untreated ultimately leads to irreversible blindness. To date, the biological mechanisms causing POAG are still unclear. A vascular hypothesis of unstable ocular perfusion has been suggested to explain the process of optic nerve damage. However, no study has previously provided a detailed exploration of the biology underlying the potential vascular connection with POAG. The Global Biobank Meta-Initiative (GBMI), a collaboration of 20 global biobanks, provides an exceptional resource to examine potential shared vascular POAG biology. A large-scale Genome-wide association studies meta-analysis was conducted in subjects sourced from 15 global biobanks and from six ancestries (n=1,487,447). A total of 59 statistically significant loci-trait associations were identified, seven of which were novel. Four loci encompassing the genes *ZFP91-CNTF, GLYAT, KALRN, CCDC13, MIR2054* and *INTU* were tested and replicated in an independent POAG cohort (n=383,500). To add biological context to these variant-trait associations, we performed TWAS using JTI cis models in 24 GTEx tissues potentially relevant to ocular conditions. We then performed fine-mapping identifying 29 gene-trait associations, five of which have been implicated or have vascular related functions: *CDKN2B, SLC35E2A, ITGB5* and *MYL4*. Further, we performed a gene enrichment analysis in which morphology and development of blood vessels, and angiogenesis pathways were significantly enriched. A gene prioritization analysis found 60 co-regulated genes of which 39 were novel. In total, 15 of the 39 novel POAG genes identified were associated with blood regulation, cardiac disease and arterial stiffness measurement. We did extensive statistical validation analysis of genes in *SIX6* and *CDKN2B-AS1* loci, previously implicated in POAG, cardiovascular diseases and cancers across multiple ancestries. Results from this analysis confirmed that the TWAS association signals in these loci are attributed to the sentinel rs33912345 missense variant in the SIX6 gene and variants linked to the *CDKN2B-AS1* gene. We also found evidence of significant interaction between the rs33912345 and causal variants in chr9p21.3, with concomitant effect on expression of the genes *CDKN2A* and *CDKN2B*. We confirmed shared biology between cardiovascular diseases and POAG by performing meta-analysis cis model TWAS-PheWAS across the whole phenome in BioVU and UKbiobank data (n=456,423). Taken together, these findings enforce the contribution of genes involved in vascular mechanisms to POAG pathogenesis.

PrgmNr 2453 - GWAS and ExWAS of blood Mitochondrial DNA copy number identifies 71 loci and highlights a potential causal role in dementia

View session detail

Author Block: **M. Chong**[1], P. Mohammadi-Shemirani[1], N. Perrot[1], W. Nelson[1], R. Morton[1], S. Narula[1], R. Lali[2], I. Khan[1], M. Khan[1], C. Judge[1], T. Machipisa[1], N. Cawte[1], M. O'Donnell[3], M. Pigeyre[1], L. Akhabir[1], G. Pare[4]; [1]Population Hlth.Res. Inst. (Genetic Molecular Epidemiology Lab), Hamilton, ON, Canada, [2]Population Hlth.Res. Inst., Hamilton, ON, Canada, [3]Natl. Univ. of Ireland Galway, Galway, Ireland, [4]Hamilton, ON, Canada

Disclosure Block: **M. Chong:** Consultant/Consulting Fees/Other Remuneration; Bayer.

Background: Mitochondrial DNA copy number (mtDNA-CN) is an accessible blood-based measurement believed to capture underlying mitochondrial function. The specific biological processes underpinning its regulation, and whether those processes are causative for disease, is an area of active investigation.

Methods: We developed a novel method for array-based mtDNA-CN estimation suitable for biobank-scale studies, called â□□AutoMitoCâ□□. We applied AutoMitoC to 395,781 UKBiobank study participants and performed genome and exome-wide association studies, identifying novel common and rare genetic determinants. Finally, we performed two-sample Mendelian Randomization to assess whether genetically low mtDNA-CN influenced select mitochondrial phenotypes.

Results: Overall, genetic analyses identified 71 loci for mtDNA-CN, which implicated several genes involved in rare mtDNA depletion disorders, dNTP metabolism, and the mitochondrial central dogma. Rare variant analysis identified SAMHD1 mutation carriers as having higher mtDNA-CN (beta=0.23 SDs; 95% CI, 0.18- 0.29; P=2.6x10-19), a potential therapeutic target for patients with mtDNA depletion disorders, but at increased risk of breast cancer (OR=1.91; 95% CI, 1.52-2.40; P=2.7x10- 8). Finally, Mendelian randomization analyses suggest a causal effect of low mtDNA-CN on dementia risk (OR=1.94 per 1 SD decrease in mtDNA-CN; 95% CI, 1.55-2.32; P=7.5x10-4).

Conclusions: Altogether, our genetic findings indicate that mtDNA-CN is a complex biomarker reflecting specific mitochondrial processes related to mtDNA regulation, and that these processes are causally related to human diseases.

PrgmNr 2454 - Higher BMI causes lower odds of depression in individuals of East Asian Ancestry

View session detail

Author Block: **J. O'Loughlin**[1], F. Casanova[2], A. Hughes[3], J. Bowden[1], E. Watkins[1], R. Freathy[1], R. G. Walters[4], L. Howe[3], K. B. Kuchenbaecker[5], J. Tyrrell[6]; [1]Univ. of Exeter, Exeter, United Kingdom, [2]Univ. of Exeter, Exeter, [3]Univ. of Bristol, Bristol, United Kingdom, [4]Univ. of Oxford, Oxford, United Kingdom, [5]Univ Coll. London, London, United Kingdom, [6]Univ. of Exeter, Exeter, Devon, United Kingdom
Disclosure Block: **J. O'Loughlin:** None.
There is extensive evidence linking higher body mass index (BMI) to higher risk of depression in adult European ancestry populations. However, our understanding of the relationship across different settings and ethnicities is limited. A recent East Asian (EAS) genome-wide association study (GWAS) of major depression showed an inverse genetic correlation of major depression with BMI (r_g=-0.212, SE=0.084). Here, we investigate the relationship between BMI and depression in EAS using a) summary statistics from the recent EAS major depression GWAS and b) individual level data in the China Kadoorie Biobank (CKB). Mendelian randomisation (MR) methods were used to test the relationships of BMI with depression. Firstly, two-sample MR was performed using major depression summary statistics from up to 15,771 cases and 178,777 controls of EAS ancestry. We repeated these analyses using a clinical measure of depression and stratifying by home location status: East Asia versus UK or USA. Secondly, one-sample MR approaches were used within CKB to investigate the relationship between higher BMI and depression in up to 100,377 individuals. Here, we tested the effects in men and women separately and compared estimates in urban versus rural dwellers.2-sample MR implied that higher BMI was associated with lower risk of depression in East Asians. For example, using a clinical depression definition and the IVW method, one-SD higher genetically-instrumented BMI was associated with lower odds of depression [OR: 0.96, 95% CI:0.93, 0.98]. Stratifying by location, there was little evidence for an inverse association for samples with East Asian ancestry living in the UK or USA, but there was for those living in East Asian countries. Consistent with this, one-sample MR in CKB also found an inverse relationship between BMI and depression [OR: 0.77, 95%CI:0.63, 0.95]. Higher BMI was consistently associated with lower odds of depression in rural dwellers in the CKB but, among urban participants, the inverse relationship was only seen for men.This study is the first formal MR analysis performed in an EAS cohort and suggests an inverse relationship between BMI and depression in this population. This is in contrast to European populations, in which genetically determined higher BMI associates with higher odds of depression. This suggests potential setting-specific causality, which is further evidenced by the lack of association between BMI genetics and depression in East Asians living in the USA or UK. Further this study highlights the importance of using ethnically diverse data, especially when the relationship under study is not purely biological and may involve socio-cultural factors.

PrgmNr 2455 - Identification of known and novel long non-coding RNAs potentially responsible for the effects of BMD GWAS loci

View session detail

Author Block: **A. Abood**[1], I. d. mesner[2], B. M. Al-Barghouthi[1], W. Rosenow[1], E. F. Morgan[3], L. Gerstenfeld[3], C. R. Farber[2]; [1]Univ. of Virginia, Charlottesville, VA, [2]Univ Virginia, Charlottesville, VA, [3]Boston Univ., Boston, MA

Disclosure Block: **A. Abood:** None.

Osteoporosis, characterized by low bone mineral density (BMD), is the most common complex disease affecting bone and constitutes a major societal health problem. Genome wide association studies (GWAS) have identified over 1100 associations influencing BMD. It has been shown that perturbations to long non-coding RNAs (lncRNAs) influence BMD and the activities of bone cells; however, the potential role of lncRNAs in explaining the genetics of BMD is unknown. Here, we applied allele-specific expression analysis, a Transcriptome Wide Association Study (TWAS), and Bayesian colocalization to RNA-seq data from human acetabular bone fragments (N=17; 5 males and 12 females; ages 43 to 80) and the Genotype-Tissue Expression (GTEx) project to identify lncRNAs potentially responsible for GWAS associations. We identified eight lncRNAs in bone RNA-seq data that are located in proximity to a BMD GWAS association and are under allele-specific expression. Using GTEx data we identified an additional 31 lncRNAs whose expression was associated (FDR0.1). The 39 lncRNAs are located in 43 BMD associations. To further support a causal role for the lncRNAs, we show that 23 of the 39 lncRNAs are differentially expressed as a function of osteoblast differentiation. Our approach identifies lncRNAs that are potentially responsible for BMD GWAS associations and suggest that lncRNAs play a role in the genetics of osteoporosis.

PrgmNr 2456 - Identifying the mediator metabolites between obesity and osteoporosis through high-dimensional mediation analysis

View session detail

Author Block: **G. Zhang**[1], A. Zhang[1], K-J. Su[2], Q. Zhao[3], Z. Wang[4], C. Mo[4], M. Brotto[4], L. Zhao[2], Z. Luo[2], Q. Tian[2], H. Shen[2], Y-P. Wang[1], H-W. Deng[2]; [1]Dept. of BioMed. Engineering, Tulane Univ., New Orleans, LA, [2]Tulane Ctr. for BioMed. Informatics and Genomics, Tulane Univ., New Orleans, LA, [3]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [4]Coll. of Nursing&Hlth.Innovation, Univ. of Texas at Arlington (UTA), Arlington, TX

Disclosure Block: **G. Zhang:** None.

Osteoporosis, mainly characterized by low bone mineral density (BMD), is a common metabolic bone disorder. Obesity has been shown positively associated with osteoporosis, and weight loss leads to bone loss. However, the molecular links between obesity and osteoporosis are still largely unknown. Investigating the biological mechanisms underlying the impact of obesity on BMD may contribute to a better understanding of bone regulation, reducing potential bone loss during weight loss in obese patients. This study aimed to identify metabolic mediators and pathways between obesity and BMD using a metabolomics and lipidomics approach. Considering the inherent correlations among the high-dimensional metabolites/lipidomics data, we built a framework with a high-dimensional multivariate mediation analysis (HDMMA) to identify the potential intermediate groups of metabolites/lipids potentially causally mediating variation from body mass index (BMI) to BMD. Using liquid chromatography-mass spectrometry-based metabolomics and lipidomics profiling, a total of 279 serum metabolites/lipids from 136 Caucasian women aged 20 to 40 in the Louisiana Osteoporosis Study (LOS) were evaluated. We performed the HDMMA for the potential metabolites group identification, and applied the causal inference using Non-combinatorial Optimization via Trace Exponential and Augmented lagRangian for Structure learning (NOTEARS) algorithm to pinpoint the critical metabolites. Finally, enrichment analysis (MetaboAnalyst 4.0) was employed to identify the metabolomic pathways associated with BMI and BMD. Five metabolitesâ□□ groups were identified as significant mediating components in which we rank the top metabolites including 3-methyl-2-oxovaleric, anthranilate, 1-aminocyclopropane-1-carboxylate, succinic acid, taurine, DHA, and 4-oxoproline. The enrichment analysis confirmed four significant pathways including tryptophan metabolism, branched-chain amino acids degradation, bile acid biosynthesis, and carnitine synthesis. Overall, our computational findings contribute considerably to understanding the novel metabolite interplay underlying the relationship between obesity and osteoporosis. Specifically, we identified several mediation pathways between BMI and BMD, which may generate metabolomic-guided hypotheses for further functional experiment studies.

PrgmNr 2457 - Incorporating the gene-gene regulatory network into the estimation of effect size

View session detail

Author Block: **Y. Zhang**, W. Dempsey, S. Zoellner; Univ. of Michigan, ANN ARBOR, MI
Disclosure Block: **Y. Zhang:** None.
Over the last few years, genome-wide association studies (GWAS) have been a successful tool to unveil genetic risk variants to a variety of complex diseases. However, additional analyses are required to identify disease susceptibility variants underlying GWAS findings and estimate their true effect sizes. For this purpose, we can leverage sources of biological information beyond analyzing the patterns of linkage disequilibrium. As most GWAS hits modify gene transcription, transcriptional data provides such biological information. However, including transcription requires modeling the interaction of gene expression in the underlying pathways of networks. In this project, we aim to incorporate gene-gene interaction network information into the posterior estimation of causal variants.
To be more specific, we aim to construct the gene expression network through gene-gene interaction data. The set of closely interacting genes in the regulatory network can be identified by existing community detection methods, such as the modularity-based or the profile likelihood-based algorithms. This will give us the transcriptional level annotation for each SNP. That is a binary matrix indicating whether the variants belong to one of the gene regulatory modules. We treat the annotation information as the prior and integrate it with the GWAS summary statistics. The posterior of the effect sizes can be achieved through the EM algorithm.
With a better understanding of the biological networks, we expect the identification of the causal variants, and the estimation of the effect sizes will be more precise. The results of our work can shed light on the biological interpretation of the direct GWAS results, as well as contribute to broad fields, such as the construction of the polygenic risk score.

PrgmNr 2458 - Mapping the proteo-genomic convergence of human diseases

View session detail

Author Block: **M. Pietzner**[1], E. Wheeler[1], J. Carrasco-Zanini[1], A. Cortes[2], M. Koprulu[1], M. WÃ¶rheide[3], E. Oerton[1], J. Cook[1], I. Stewart[4], N. Kerrison[1], J. Luan[1], J. Raffler[3], M. Arnold[3], W. Arlt[5], S. O'Rahilly[1], G. KastenmÃ¼ller[3], E. Gamazon[6], A. Hingorani[7], R. A. Scott[8], N. Wareham[1], C. Langenberg[9]; [1]Univ. of Cambridge, Cambridge, United Kingdom, [2]Stevenage, United Kingdom, [3]Helmholtz Zentrum MÃ¼nchen, Neuherberg, Germany, [4]MRC Epidemiology Unit, Cambridge, United Kingdom, [5]Univ. of Birmingham, Birmingham, United Kingdom, [6]Vanderbilt Univ. Med. Ctr., Nashville, TN, [7]Univ. Coll. London, London, United Kingdom, [8]GSK, Stevenage, United Kingdom, [9]Univ Cambridge, Cambridge, Cambridgeshire, United Kingdom

Disclosure Block: **M. Pietzner:** None.

Characterization of the genetic regulation of protein abundance and function is essential for understanding disease aetiology and identifying new therapies. Here we identify 10,674 genetic associations for 3,892 plasma proteins to create the first *cis*-anchored gene-protein-disease map of 1,859 connections that highlights strong cross-disease biological convergence. For example, the genetic signal at *EFEMP1* for FBLN3 was shared across diverse connective tissue disorders consistent with abnormal elastic fibre morphology *in the Efemp1* knock-out mice. Integration of diverse â□□omicâ□□ layers identifies a supersaturated bile to promote cholesterol crystallization and gallstone formation as mode of action of *SULT2A1*. We demonstrate the specific value of *cis*-protein quantitative trait loci (pQTLs) for causal annotation of disease genes at established GWAS loci, such as *PRSS8* underlying the Alzheimerâ□□s signal currently assigned to *KAT8.* We develop a new data-driven methodology that identifies 39% of 2,303 distant protein-associated *trans*-pQTLs as protein- or pathway-specific through their protein-network profile, including known disease variants such as for *PNPLA3* and a community of metabolic and detoxification enzymes. Our results establish proteo-genomic connections with and among diseases even of seemingly unrelated aetiology and help to address one of the major barriers for experimental validation and clinical translation of GWAS discoveries. We make results accessible via an interactive web resource upon publication (https://www.omicscience.org/).

PrgmNr 2459 - Mitochondrial and sex chromosome genetically regulated gene expression implicates new genes in complex traits across multiple human populations

View session detail

Author Block: **D. AraÃºjo**[1], S. S. Rich[2], J. I. Rotter[3], H. Im[4], A. W. Manichaikul[2], H. E. Wheeler[5], NHLBI TOPMed Consortium; [1]Loyola Univ. Chicago, Chicago, IL, [2]Univ Virginia, Charlottesville, VA, [3]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [4]The Univ. of Chicago, Chicago, IL, [5]Loyola Univ Chicago, Chicago, IL

Disclosure Block: **D. AraÃºjo:** None.

The majority of GWAS are conducted in European ancestry populations and are limited to autosomal chromosomes, ignoring the genetic content of the mitochondria and sex chromosomes. Alongside GWAS, transcriptome-wide association studies (TWAS) can provide useful information about the direction of gene regulation underlying complex traits. Given the genetic diversity among individuals, we sought to build mitochondrial and sex chromosome transcriptome prediction models for use in TWAS in diverse populations, including those underrepresented in GWAS and TWAS.

We used transcriptome data from the Multi-Ethnic Study of Atherosclerosis (MESA) comprised of up to 1004 individuals of African, Chinese, European and Hispanic/Latino ancestries. For each of 3 blood cell types, peripheral blood mononuclear cells (PBMC), CD16+ monocytes, and CD4+ T cells, we built models in each population and also a model including all individuals. We used cross-validated elastic net to estimate gene expression from local SNPs within 1Mb of each gene through an additive linear model. Depending on population, our modeling resulted in 24-57 genes with Spearman correlation Ï>0.1. Smaller sample sizes were available for monocytes and T cells, resulting in 3-16 and 4-13 genes with Ï>0.1, respectively. Most predicted genes were on the X chromosome, while few Y chromosome and mitochondrial genes had Ï>0.1.

With these prediction models, we applied S-PrediXcan to X chromosome GWAS summary statistics from two different multi-ancestry studies, the Population Architecture using Genomics and Epidemiology (PAGE) study (n=49,839) and Pan UK Biobank (PanUKB, n=488,377). We identified 5 gene-trait pairs that were significant in both studies (PGRIPAP1 associated with diastolic blood pressure, *STARD8* with platelet count, *PLXNA3* with triglyceride levels, and both *TSC22D3* and *SPIN2B* with height. Of these 5 gene-trait pairs, only *STARD8* - platelet count association had been reported previously; thus, the remaining 4 correlations may be novel. For the mitochondrial genes, GWAS summary statistics were only available from the UK Biobank. We identified statistically significant correlations between *MT-ND3* and mean corpuscular hemoglobin, mean corpuscular volume, mean platelet volume, plateletcrit, red blood cell count and red blood cell width distribution (P-7). We expect that conducting more integrative omics studies that include mitochondria and sex chromosomes in multi-ethnic cohorts will identify new gene-trait associations and promote diversity in biomedical research.

PrgmNr 2460 - Parent-of-origin inference in the UK biobank reveals new imprinted variants

View session detail

Author Block: **R. Hofmeister**[1,2], S. Rubinacci[1,2], D. Ribeiro[1,2], Z. Kutalik[3,2], A. Buil[4], O. Delaneau[1,2]; [1]Dept. of Computational Biology, Univ. of Lausanne, Lausanne, Switzerland, [2]Swiss Inst. of Bioinformatics (SIB), Lausanne, Switzerland, [3]Unisante, Lausanne, Switzerland, [4]Mental Hlth.Sct Hans Hosp., Roskilde, Denmark
Disclosure Block: **R. Hofmeister:** None.
Sex-specific genetic effects have been reported in many diseases and phenotypes through Genome-Wide Association Studies (GWAS) conducted separately on men and women. Recently, GWAS have pinpointed more complex sex-specific effects in which identical genetic variations have different phenotypic effects depending on their parent-of-origin (PO). Yet, such studies were largely limited in terms of sample size: researchers usually rely on the parental genomes or known genealogies to characterize the PO. However, these requirements are rarely available in biobanks, preventing the study of PO effects on a large number of phenotypes and samples.
Here, we present a novel probabilistic approach to determine the PO of an individualâ□□s haplotypes that does not require prior knowledge of genealogy. Our model (i) identifies surrogate parents using identity-by-descent (IBD) sharing between individuals, (ii) uses a Hidden Markov Model to represent a specific haplotype as a mosaic of haplotypes of its close relatives, (iii) assigns paternal or maternal origin to these shared haplotype segments, and (iv) extends this assignment to entire chromosomes. Using the UK Biobank dataset, we were able to infer the PO for ~25,000 samples. This represents a five-times increase of sample size compared to classical approaches. We assessed the accuracy of our inference using a total of 1,397 UK biobank duos and trios, resulting in a call rate of ~75% (% of sites with PO assignment) and an error rate of ~0.5% (% heterozygous sites with incorrect PO assignment). We then performed association scans for PO effects across hundreds of phenotypes, ranging from physiological traits to complex diseases, allowing us to replicate known and discover new imprinted variants at a scale never reached before. We also showed that many of these variants could reach genome-wide significance only thanks to the larger sample size offered by our approach. As an example, we discovered four loci associated with platelet crit exhibiting strong PO effects. One of these loci corresponds to an expression Quantitative Trait Loci of the well-known imprinted gene MEG3 (Maternally Expressed Gene 3), previously shown to be involved in the platelet phagocytosis. Our approach, designed to leverage the relatedness of individuals in large biobanks, greatly increases the effective sample size for PO studies and therefore offers the means necessary to further characterize the genetic architecture of sex-specific effects in complex traits and diseases.

PrgmNr 2461 - Phenome-wide association study of clonal haematopoiesis somatic mutations in 391,756 UK Biobank participants

View session detail

Author Block: **J. Mitchell**, K. J. Carss, Q. Wang, A. R. Harper, S. Petrovski; Ctr. for Genomics Res., Discovery Sci., R&D, AstraZeneca, Cambridge, United Kingdom
Disclosure Block: **J. Mitchell:** Salary/Employment; AstraZeneca.
Clonal haematopoiesis (CH) is a common, age-related process that describes the clonal expansion of blood cells with somatic mutations (i.e. variant allele frequency (VAF) >2%), predominantly in three genes (*DNMT3A*, *T*ET2, *ASXL1*). In addition to being a well-established risk factor for haematological malignancy, CH has demonstrated strong associations with cardiovascular disease and all-cause mortality. To further appreciate gene-phenotype associations attributable to CH we analysed exome sequencing data from 391,756 UK Biobank participants. Our approach repurposed a germline-optimised calling pipeline to enable the detection of somatic variants with high VAF. Adoption of this approach identified protein truncating (putatively somatic) variants across *DNMT3A*, *TET2* and *ASXL1* in 2,888 individuals with a mean VAF of 0.276 Â± 0.093. This is higher than a recent Trans-Omics for Precision Medicine program study which, utilising a somatic variant calling pipeline, reported a CH variant mean VAF of 0.190 Â± 0.102 and largely accounts for the approximately four times lower detectable CH rate observed in the study presented here. As expected, and comparable with other large sequencing based efforts characterising CH, prevalence was correlated with age ($\tilde{A}$ = 0.096, $P = 1.91x10^{-190}$). To discover novel trait associations, and further investigate known associations, we performed a gene-level collapsing analysis-based phenome-wide association study (PheWAS) of the detectable somatic mutations with 16,259 binary and 1,568 quantitative phenotypes. We found 251 statistically significant (*P* -8) binary trait associations, of which 81 (32%) were not cancer-related, and 33 associations with quantitative traits. Of the three studied genes, most associations were with *ASXL1* (n=142) with slightly fewer for *TET2* (n=135) and far fewer for *DNMT3A* (n=7). We have analysed sequencing data for CH mutations in the largest cohort to date and validated and expanded upon their known trait associations. We will repeat these analyses adopting the calls from a somatic-optimised bioinformatics pipeline to increase our sensitivity for CH detection in these three genes.

PrgmNr 2463 - Shared genetic effects for type 2 diabetes and four cancers uncovered through multi-phenotype genome-wide association study

View session detail

Author Block: **A. Demirkan**[1], L. Zudina[1], I. Pupko[1], Z. Balkhiyarova[1], A. Ulrich[1], J. Maina[2], P. Froguel[3], E. Riboli[4], M. J. Gunter[5], M. Kaakinen[1], I. Prokopenko[1]; [1]Univ. of Surrey, Guildford, United Kingdom, [2]CNRS UMR8199 EGID (European Genomic Inst. for Diabetes), Lille, France, [3]CNRS UMR - 8090, Lille Cedex, France, [4]Inst. Pasteur de Lille, Univ. of Lille, Lille, France, [5]Nutrition and Metabolism Section, Intl. Agency for Res. on Cancer, World Hlth.Organization, Lyon, France
Disclosure Block: **A. Demirkan:** None.
Introduction: There are established relationships between type 2 diabetes (T2D) and cancer, including cancer being the most common cause of death in T2D. We aimed to gain insights into the pathophysiological processes shared between T2D and four cancers through multi-phenotype (MP) genome-wide association study (GWAS). Methods: We combined GWAS on 36,173 individuals from the pan-European EPIC study, including 10,855 T2D cases, 4,126 postmenopausal breast, 2,111 colorectal, 473 pancreatic and 419 prostate cancer cases. The combined GWAS dataset was quality controlled and imputed against the HRC reference panel providing 39.3M DNA variants for analysis. We performed MP-GWAS reverse regression for five outcomes using SCOPA software. We evaluated single and multi-phenotypes effects at associated (P10 as important model fit improvement. Results: Within MP-GWAS, we identified 174 independent loci for the full five-phenotype model. Among them, 31 overlapped established T2D/cancer variants, specifically we replicated 12 loci associated with breast cancer, 10 with prostate, one with colorectal cancer, and 15 with T2D. Seven of the previously established loci were reported for multiple cancer/T2D outcomes before. These were rs67798996 (ASCL2), rs57096576 (SMIM38), rs12524664 (TNXB), rs2019689 (MBNL1) rs7756992 (CDKAL1), rs35011184 (TCF7L2) and rs62033406 (FTO), each previously reported for two out of five outcomes we studied. Six of these (ASCL2, SMIM38, MBNL1, TNXB, CDKAL1, TCF7L2 and FTO), except TNXB yielded better model fit for single phenotype tests as defined by deltaBIC. The 143 novel signals included 112 performing better in single-phenotype association, 10 loci with the five-phenotype model suggesting T2D-cancers shared effects. Conclusions: The large data and power, improved through implementation of multi-variate GWAS analysis enabled identification of 10% of loci with shared T2D-cancer effects that contribute to these diseasesâ comorbidity through shared genetic effects. Funding: WCRF-2017/1641, LonglTools H2020-SC1-2019-874739, PreciDIAB, ANR-18-IBHU-0001

PrgmNr 2464 - The chromatin accessibility signature of selection for exercise capacity in skeletal muscle

View session detail

Author Block: **P. Orchard**[1], N. Manickam[1], M. K. Treutelaar[1], Y. Zhang[1,2], C. R. Evans[1], J. Z. Li[1], C. F. Burant[1], S. C. J. Parker[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Soochow Univ., Suzhou, China
Disclosure Block: **P. Orchard:** None.
Aerobic exercise capacity is negatively associated with mortality and complex disease susceptibility. Understanding the genetics and epigenomics of aerobic exercise capacity may clarify the molecular mechanisms behind this relationship. Artificial selection on untrained running capacity was previously used to generate two rat lines, one with high capacity runners (HCRs) and one with low capacity runners (LCRs). The lines show >9-fold difference in running capacity, and heritability estimates of running capacity within each line are >0.4; however, a genome wide association study on running capacity in a HCR-LCR F2 population identified no genome-wide significant associations, suggesting the trait is highly polygenic. To understand the noncoding landscape driving differences in exercise capacity and the response to exercise, we profiled chromatin accessibility (ATAC-seq) on 253 rat skeletal muscle samples, representing HCRs and LCRs before, during, and 1 hour after exercise, and from a HCR-LCR F2 cross.

We identified 4,932 HCR-LCR differential open chromatin regions (5% FDR). Consistent with the hypothesis that the cross-line chromatin accessibility differences are driven by genetics and are selected for, we find cross-line genetic differences (measured by F_{ST}) are significantly enriched ($p = 9.1x10^{-16}$) in HCR-LCR differential peaks but not differential peaks from other comparisons. Peaks with greater accessibility in HCR rats (HCR-up peaks) are enriched ($p = 1.2x10^{-8}$) for nuclear receptor family transcription factor (TF) motifs, nominating their role in determining exercise capacity. HCR-up peaks showed inflated p-values for positive correlation with running capacity in the F2 rats, suggesting that the differential peaks contribute to the cross-line difference in running capacity. We performed a chromatin accessibility quantitative trait locus (caQTL) scan using the F2 rats and identified 7,775 significant peaks (5% FDR). HCR-LCR differential peaks were more likely than non-differential peaks to associate with a caQTL (p -300), indicating SNPs in the F2 population that may drive the line-differential peaks.

We identified 22,731 differential peaks between rats at rest vs. during exercise, and 32,095 differential peaks between rats at rest vs. 1 hour post exercise (5% FDR). We find MEF2 and ATF3 TF motifs are enriched in exercise-responsive peaks, and a motif for NRF2, a regulator of the antioxidant response, is enriched in peaks with greater accessibility post exercise.

These results highlight the widespread epigenomic response to exercise and nominate transcription factors that underlie exercise capacity and response to exercise.

PrgmNr 2465 - The effect of mitochondrial population genome variation on nuclear gene expression and its link to disease

View session detail

Author Block: **J. Lascano Maillard**[1,2], H. Ongen[3,2], E. T. Dermitzakis[4,2]; [1]Genetic Med. and Dev., Univ. of Geneva, Geneva, Switzerland., Geneva, Switzerland, [2]Swiss Inst. of Bioinformatics, Geneva, Switzerland, [3]Genetic Med. and Dev., Univ. of Geneva, Geneva, Switzerland, Geneva, Geneva, Switzerland, [4]Univ. of Geneva, Geneva, Switzerland
Disclosure Block: **J. Lascano Maillard:** None.
Mitochondria are involved in different key aspects of cell homeostasis. Somatic mutations in the mitochondrial genome are linked to varying patterns of nuclear DNA methylation and acetylation, contributing to differential gene expression and to the development of particular diseases such as diabetes, encephalomyelitis and cardiomyopathies. Although mitochondrial sequence alteration was investigated in the pathological context, little is known about the effect of its population variation on phenotypic traits. Two previous studies on whole blood containing a mix of cell types examined the impact of variation among northern European and the Finnish population, respectively, finding only a few mito-nuclear expression quantitative trait-loci (eQTLs). Here, we extended the analysis using 358 samples of four European regions from the *1000 Genomes* consortium and matched lymphoblastoid cell lines (LCLs) RNA-seq data from the *GEUVADIS* project giving us the potential to unravel more mitochondrial genetic associations with nuclear gene expression. First, to assess the extent of genetic control of the mitochondrial genome on nuclear genes, we performed an eQTL analysis using mitochondrial genotypes and nuclear gene expression data. We further explored potential gene pathways involving mito-nuclear eQTL genes using Bayesian Networks. We find a total of 66 mito-nuclear eQTLs involving 21 different mitochondrial variants and 65 unique nuclear genes. Mitochondrial eQTL variants were grouped into high LD genotypic blocks ($r2 > 0.7$) and the corresponding eQTL genes were enriched for ontology terms related to mitochondrial genetic diseases, such as encephalomyelitis and muscular dystrophy. Each genotypic block defining specific mitochondrial haplogroups were associated with eQTL genes enriched in the corresponding disease terms from known specific haplogroup-enriched diseases, such as osteoarthritis and auditory-loss. Within genotypic blocks, eQTL genes were connected within pathways and included the known *FOXP2-TDO2* axis. In conclusion, we show that the mitochondrion carries genetic determinants for nuclear gene expression that might explain predisposition to specific diseases. Our findings point towards an expanded role of the mitochondrial genome on human phenotypic variation and pave the way for future studies which would include mitochondrial SNPs in polygenic risk scores calculation.

PrgmNr 2466 - The Netherlands Neurogenomics Database - Genetic susceptibility for brain disorders, clinical disease trajectories and end-stage neuropathology

View session detail

Author Block: **I. Holtman**[1], E. Boddeke[1], B. Eggen[1], J. Hamann[2], M. Swertz[1], A. Rozemuller[2], S. Wehrens[3], M. Groot[3], I. Huitinga[3]; [1]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [2]Amsterdam Med. Ctr., Amsterdam, Netherlands, [3]Netherlands Inst. for NeuroSci., Amsterdam, Netherlands
Disclosure Block: **I. Holtman:** None.
The brain is a highly complex organ that consists of many intricately linked substructures and cell types and is susceptible to a wide range of psychiatric and neurodegenerative diseases. The Netherlands Brain Bank (NBB) has performed more than 4.500 human brain autopsies from donors with a wide range of psychiatric and neurodegenerative conditions as well as matched controls. From each brain, brain regions were dissected according to disease-specific protocols, and a range of clinical and pathological data were collected in extensive clinical-pathological summaries. We recently initiated the Netherlands Neurogenomics Database, which aims to integrate the extensive clinical and neuropathological data, with a multi-omics map from a large number of donors in order to study the effects of genetic variation on brain disease. Using text mining approaches, we aimed to convert the vast amount of qualitative clinical neuropathological summaries into standardized clinical and neuropathological trait assessments. We identified a set of 80 key clinical and neurological parameters and manually labeled almost 20.000 sentences. This data was used to train and compare different Google Bidirectional Encoder Representations from Transformers (BERT) multilabel classification models. The trained model was highly reliably able to score these predefined parameters. As these clinical parameters have a temporal aspect, we are currently implementing graph-based approaches to convert these data types into clinical disease trajectories. Neuropathological examination reports are converted into standardized neuropathological trait assessments using a combination of BERT based multilabel classification and regression approaches. Additionally, we are processing Illumina GSA SNP chips for around 3000 donors and aim to calculate polygenic risk scores (PRS) for a wide range of Central Nervous System (CNS) disorders and neurobehavioral traits. These PRS will be used to determine the relationship between genetic susceptibility and clinical disease trajectories and end-stage neuropathology. This unique resource will be open-access and made readily available to the large scientific community. It will facilitate many existing lines of research into the understanding of CNS disease susceptibility.

PrgmNr 2467 - Trio clinical exome analysis reveals a digenic phenotype in a 19-months-old boy with congenital deafness

View session detail

Author Block: R. Zeuli[1], **C. Peduto**[1], M. Zanobio[1], F. Romano[1], M. E. Onore[1], G. Blasio[1], A. Torella[1,2], G. Cappuccio[3,2], F. Del Vecchio Blanco[1], G. Piluso[1], N. Brunetti Pierri[3,2], V. Nigro[1,2]; [1]Dept. of Precision Med., Univ. of Campania â□□Luigi Vanvitelliâ□□, Naples, Italy, [2]Telethon Inst. of Genetics and Med., Pozzuoli, Italy, [3]Dept. of Translational Med., Univ. of Naples â□□Federico IIâ□□, Naples, Italy
Disclosure Block: **C. Peduto:** None.
Background: In most laboratories multigene NGS panels are used to diagnose specific phenotypes. However, in infants and young children some conditions may be missed using focused gene analysis. Trio clinical exome can accelerate the discovery of additional genetic variants not related to the referring phenotype of the patient but can still have clinical implications. Here, we present the case of a child with congenital deafness.
Patients concerns: The family came to our observation for genetic counselling about the familiar congenital deafness of a 19-month-old baby. While the NGS analysis was in progress, the patient was found with elevated serum creatine kinase levels and mild development delay.
Results: We detected two compound heterozygous missense variants in the TBC1D24 gene causing deafness. Strikingly, we found an additional maternally inherited variant in the intron 55 of DMD gene c.8217+1G>T. This allele is new but predicted to cause Duchenne Muscular Dystrophy through an out-of-frame skipping of exon 55. The genetic diagnosis was done prior of any clinical sign of muscular dystrophy.
Conclusions: Here we report an example of the urgency of trio clinical exome analysis in paediatric contest instead of panels. In this case, early diagnosis was crucial in DMD to organize the best long-term care and to provide genetic counselling for the whole family.

PrgmNr 2468 - Validation of cross-ancestry and cross-trait polygenic risk scores in the Genomics England cohort

View session detail

Author Block: **S. Selzam**, F. Riveros-Mckay, D. Thompson, R. Moore, D. Wells, D. Palmer, J. Harrison, Genomics England Research Consortium, V. Plagnol, G. McVean, P. Donnelly, M. E. Weale; Genomics plc, Oxford, United Kingdom
Disclosure Block: **S. Selzam:** Salary/Employment; Genomics plc.
PRS models are often trained and evaluated within the same cohort by artificially splitting into training and testing samples. External validation is considered a more robust test of true performance. We developed PRS models for ten common diseases: four cancers (ovarian, prostate, bowel, breast), type 2 diabetes, atrial fibrillation, ischemic stroke, hypertension, coronary artery disease, and cardiovascular disease. We included a subset of the UK Biobank (UKB) data for training (N=~340,000) and investigated their predictive performance in both the held-out UKB samples and a separate UK cohort, the 100,000 Genomes Project (Genomics England; n=40,001). We boosted cross-ancestry performance and trait prediction by incorporating ancestry-specific effect sizes and linkage disequilibrium, and by leveraging cross-trait correlations. Validation of PRS performance in UKB and the 100,000 Genomes Project showed high and consistent performance, with 90% of the prediction estimates showing overlapping 95% AUC confidence intervals between the two cohorts across the different traits and ancestries. The prostate cancer PRS was our most predictive PRS model, with an AUC of 0.72 and 0.71 in UKB and the 100,000 Genomes Project, respectively. For type 2 diabetes, the cross-ancestry method improved prediction (relative AUC increase) in all ancestry groups, ranging from a 9.7% improvement in East Asian ancestries (AUC=0.68 vs. 0.62) to a 1.4% improvement in African ancestries (AUC=0.61 vs. 0.60). Our cross-trait methodology increased ischemic stroke prediction by 4.4% (AUC=0.60 vs 0.58). Our PRS models outperformed best-in-class published PRS across all traits and across a range of ancestries (mean Î AUC=0.04; mean Î improvement=7.2%), with the largest gains observed for type 2 diabetes in East Asian populations (Î AUC=0.12; prediction improvement=20.7%). Using the UK Biobank sample to train the PRS for evaluation in the 100,000 Genomes Project generally increased performance, although in some cases only by moderate amounts. The increase in prediction was largest for hypertension in South Asian ancestries, with an improvement of 6.7% (Î AUC=0.04). Our results show that our PRS models perform well in an external cohort with different study characteristics to the UKB training cohort. The calculated PRS values in individuals from UK Biobank and the 100,000 Genomes Project will be made available as an additional resource to the scientific community.This research was made possible through access to the data and findings generated by the 100,000 Genomes Project; http://www.genomicsengland.co.uk, and the UK Biobank Project, https://www.ukbiobank.ac.uk/.

PrgmNr 2469 - Whole genome sequencing association study among 24,870 individuals from multiple ancestry-ethnicity groups identifies common and rare variants associated with serum urate

View session detail

Author Block: **A. Tin**[1,2], P. Sekula[3], J. A. Brody[4], D. Jain[5], A. Hoppmann[3], L. M. Raffield[6], S-J. Hwang[7], R. Irvin[8], W. Zhao[9], I. Chen[10], J. Blangero[11], J. S. Floyd[5], R. Loos[12], P. F. McArdle[13], E. Boerwinkle[14], H. Kramer[15], N. Franceschini[16], A. Kottgen[17,2], NHLBI TOPMed Consortium; [1]Univ. of Mississippi Med. Ctr., Jackson, MS, [2]Johns Hopkins Univ., Baltimore, MD, [3]Univ. of Freiburg, Freiburg, Germany, [4]Univ of Washington, Seattle, WA, [5]Seattle, WA, [6]UNC - Chapel Hill, Chapel Hill, NC, [7]Natl. Inst. of Hlth., Framingham, MA, [8]Univ. of Alabama, Tuscaloosa, AL, [9]Univ California los Angeles, Los Angeles, CA, [10]Cedars Sinai Med Ctr, Los Angeles, CA, [11]Univ. of Texas Rio Grande Valley Sch. of Med., Brownsville, TX, [12]The Icahn Sch. of Med. at Mount Sinai, New York, NY, [13]Univ Maryland Sch Med, Baltimore, MD, [14]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [15]Loyola Univ. Med. Ctr., Maywood, IL, [16]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [17]Univ Hosp Freiburg, Freiburg, Germany
Disclosure Block: **A. Tin:** None.
Background Elevation of serum urate levels can cause gout, a common and painful form of inflammatory arthritis. Previous genome-wide association studies (GWAS) of serum urate based on genotyped and imputed variants have identified nearly 200 genetic loci. The index variants at most loci are common and map outside of coding regions. Whole genome sequencing (WGS) provides more comprehensive coverage of the genome and provides the opportunity to identify rare variant associations in both coding and non-coding regions. **Methods** We performed a transethnic WGS association study of serum urate among participants of the Trans-omics for Precision Medicine (TOPMed) program using the Analysis Commons platform. After data cleaning, serum urate levels were transformed to be normally distributed. A kinship matrix was used to account for relatedness. The analysis controlled for age, sex, body mass index, cohort-ancestry-ethnicity strata, and genetic principal components. Single variant analysis (SVA) was conducted using the GENESIS (GENetic EStimation and Inference in Structured samples) software and included variants with minor allele count â¥10. Gene-centric aggregated variant analysis was conducted using STAAR (variant-Set Test for Association using Annotation information). Aggregation analysis of variants with minor allele frequency (MAF) Results We included 24,870 participants from multiple ancestry-ethnicity groups from 10 cohorts. The genomic control factor of SVA of serum urate indicated little inflation (1.03). After successively conditioning on the most significant variant(s) in each chromosome, we identified 12 independent single variant associations (4 with MAF SLC2A9 and *SLC22A12* encoding urate transporter genes contained 3 and 4 independent variants, respectively. Aggregation analysis identified significant rare variant associations with lower urate levels at *SLC22A12* (median effect size: -1.4 SD), driven by disruptive missense and loss-of-function (LoF) variants as well as variants in promoter and upstream regulatory regions. Rare variant signals at *SLC2A9* was driven by disruptive missense and LoF variants associated with urate in both directions. **Conclusion** This large-scale WGS study of serum urate identified potentially novel rare variants associated with serum urate levels.

PrgmNr 2470 - Whole-exome association analyses of polygenic deviators for height in 183,948 individuals from the UK Biobank

View session detail

Author Block: **G. Hawkes**[1], L. Yengo[2], S. Vedantam[3], E. Marouli[4], G. Lettre[5], Y. Okada[6], J. N. Hirschhorn[7], T. Frayling[1], A. R. Wood[1], GIANT Consortium; [1]Genetics of Complex Traits, Coll. of Med. and Hlth., Univ. of Exeter, Exeter, Devon, United Kingdom, [2]Inst. for Molecular BioSci., The Univ. of Queensland, Brisbane, Australia, [3]Endocrinology, Boston Children's Hosp, Sharon, MA, [4]William Harvey Res. Inst., Barts and The London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom, [5]Montreal Heart Inst, Montreal, QC, Canada, [6]Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., Suita, Osaka, Japan, [7]Boston Children's Hosp./Broad Inst., Boston, MA
Disclosure Block: **G. Hawkes:** None.

Genome-wide association studies have explained ~25% of normal variation in height through thousands of common genetic variants. Individuals with short stature may carry a large number of height-lowering alleles. Alternatively, they may deviate from this model because of a rare and highly deleterious variant, monogenic cause, or environmental factors. Using 183,948 individuals from the UK Biobank of European ancestry with imputed and whole-exome sequence (WES) data, we aimed to 1) define individuals as polygenic â□□deviatorsâ□□ based on height and polygenic risk scores (PRS), and 2) perform single-variant and burden association testing of rare variants to identify putative novel loci for â□□deviatorâ□□ case status.

A polygenic risk score (PRS) was generated in UK Biobank using effect estimates of 3,198 independent SNPs associated with height (P-8) available from an interim GIANT meta-analysis of 1,400,860 individuals of European ancestry that excluded UK Biobank. The correlation between height (adjusted for age, sex, and centre) and PRS was assessed relative to a simulated population with an additive genetic architecture to identify â□□shortâ□□ deviators. Deviators were defined using both Mahalanobis distances (MD) (*P*The PRS explained 32% of the variance in height within the UK Biobank samples. We identified 813 and 4,249 â□□shortâ□□ deviators based on MD and RO, respectively. Single variant association analysis identified a total of 6 missense variant associations and 1 LoF association at exome-wide significance (P-7). Putative novel associations included a missense-variant association in *NBPF4* (P=1x10^{-10}) and the LoF variant association was located in *MSH4* (P=9x10^{-9}). Burden testing identified 5 genes associated with short deviator status, all previously associated with single gene growth disorders.

In conclusion, we have identified putative novel genes by identifying individuals with phenotypes deviating from polygenic expectation.

PrgmNr 2471 - *PAX6*-associated microphthalmia: Transcriptome-wide investigation of patient-derived optic vesicle iPSC models reveals molecular disruption during early eye development associated with missense *PAX6* variant

View session detail

Author Block: **P. Harding**[1], D. Lima Cunha[1], N. Owen[1], J. Eintracht[1], M. Moosajee[1,2,3,4]; [1]UCL Inst. of Ophthalmology, London, United Kingdom, [2]The Francis Crick Inst., London, United Kingdom, [3]Moorfields Eye Hosp. NHS Fndn. Trust, London, United Kingdom, [4]Great Ormond Street Hosp. for Children NHS Trust, London, United Kingdom
Disclosure Block: **P. Harding:** None.
PAX6 is an essential transcription factor for eye development, regulating diverse genetic pathways. Heterozygous pathogenic *PAX6* variants cause a wide range of ocular disorders, including aniridia, cataracts and Peterâ□□s anomaly. Missense variants are typically associated with milder phenotypes, however, substitutions in the highly conserved DNA-binding paired domain can give rise to more severe ocular features including microphthalmia, a congenital, structural disorder resulting in a small, underdeveloped eye. Microphthalmia is reported in 11% of blind children, yet currently no treatments are available to improve visual function. The disease-causing downstream pathways affected in these patients remain unclear. We aim to clarify direct and indirect targets of *PAX6* in eye development that are perturbed in *PAX6*-associated microphthalmia.
3D optic vesicles were differentiated from iPSCs of a patient exhibiting severe microphthalmia, aniridia and cataracts diagnosed with heterozygous *PAX6* variant c.372C>A, p.Asn124Lys. Molecular characterisation was performed at 20- and 35-days differentiation through qRT-PCR and immunostaining to confirm an early-stage disease phenotype. Total RNAseq transcriptome-wide profiling was utilised to characterise global gene expression changes associated with the missense variant compared to unaffected wildtype (WT) controls and vesicles from a non-microphthalmic aniridia patient with nonsense *PAX6* variant c.781C>T, p.Arg261*.
PAX6 microphthalmic vesicles showed significantly reduced mRNA expression of early eye transcription factors *PAX6, RAX, OTX2* and *SOX2* relative to WT controls, alongside absence of neural retina marker VSX2 protein, which was present in aniridia and WT vesicles at day 35. Principal component analysis of RNAseq data displayed robust clustering of samples by disease state, with differential gene expression analysis revealing significant molecular changes, including pathways relating to cellular metabolism and growth.
We present a cellular model of *PAX6*-related microphthalmia which mimics the severe phenotype observed in patients, with early optic vesicles failing to induce eyefield transcription factors and subsequently form presumptive neural retina. Variant-specific molecular disruption identified by RNAseq provides insights into genotype-phenotype correlations observed in *PAX6* patient cohorts. This work creates a valuable resource for exploring the roles of *PAX6* in human eye development, and further investigation will enhance understanding of microphthalmia pathogenesis, which could lead to development of novel therapies.

PrgmNr 2472 - A Novel Method of Cell-type-level EWAS of HIV Infection Following Deconvoluting DNA Methylation in Blood

View session detail

Author Block: **X. Zhang**[1], A. Justice[1], B. E. Aouizerat[2], X. Ke[1]; [1]Yale Sch. of Med., New Haven, CT, [2]Univ California San Francisco, San Francisco, CA
Disclosure Block: **X. Zhang:** None.
Background. Epigenome-wide association studies (EWAS) have previously linked DNA methylation in blood to HIV infection. However, the interpretation of epigenome data from blood samples is complicated by current approaches to addressing cell composition heterogeneity. Clarifying the precise DNA methylation mechanism of HIV infection in each cell type would require the dissection of HIV infection-associated methylation signals. Here, we developed a new method to carry out cell type level EWAS for HIV infection following computational deconvolution of the methylome in each cell type.**Methods**. DNA methylation for 718 whole blood DNA samples from Veterans Aging Cohort Study (VACS) cohort was profiled by using Illumina HumanMethylation 450K Beadchip (HIV-positive=614; HIV-negative=104). Cell type proportions for each sample were estimated. We applied a Tensor Composition Analysis (TCA) approach to deconvolute cell-type-specific DNA methylation signals in 6 major cell types (CD4+ T cell, CD8+ T cell, monocyte, B cell, Nature Killer, granulocytes). We followed a two-step general linear regression model to adjust systematic bias, known biological confounders and unknown global confounders. CpGs associated with HIV infection in each cell type were declared significant at a Bonferroni adjusted p Results. We identified significant CpG sites for HIV infection in each cell type: CD4+ (107 CpGs), CD8+ (12 CpGs), B cells (2 CpGs), Nature Killer (12 CpGs), monocytes (6 CpGs), and granulocytes (8 CpGs). We replicated our previously reported HIV-associated CpG sites from bulk methylome EWAS, *NLRC5* and *LPCAT1*, which were less methylated in HIV-positive samples. Importantly, we dissected these significant signals from bulk methylome to specific cell types. Two previously identified *NLRC5* promoter region CpG sites (cg07839457, cg16411857) were hypomethylated in CD4+ T cells (t=-6.24, p_{adj}=3.34E-04; t=-6.63, p_{adj} =3.01E-05) and granulocytes (t=-7.08, p_{adj}=1.56E-06; t=-6.04; p_{adj}=1.07E-03). A third *NLRC5* CpG, cg05757530, was only significant in CD4+ T cells (t=-6.70, p_{adj}=1.35E-06). One *LPCAT1* CpG, cg16272981, was hypomethylated in CD4+ T cells, B cells, monocytes, and granulocytes (t=-7.84~-5.97, p_{adj}=1E-04~8.49E-09). Additionally, we identified a few novel CpG sites for HIV infection in different cell types. For example, we found two differentially methylated *SHANK2* CpG sites in CD4 T cells. **Conclusions**: Computationally deconvoluted DNA methylation is a robust and effective approach to dissect cell type-specific DNA methylation signals for HIV infection. Our results further reveal cellular epigenetic mechanisms involving HIV pathology.

PrgmNr 2473 - Association of peripheral blood microRNAs with cardiovascular endophenotypes in a population-based study

View session detail

Author Block: **V. Talevi**[1], G. Pehlivan[1], T. Pena Centeno[2], S. Burkhardt[2], A. Fischer[2,3], M. M.B. Breteler[1,4]; [1]Population Hlth.Sci., German Ctr. for Neurodegenerative Diseases (DZNE), Bonn, Germany, [2]Dept. of Systems Med. and Epigenetics, German Ctr. for Neurodegenerative Diseases (DZNE), Bonn, Germany, [3]Dept. of Psychiatry and Psychotherapy, Univ. Med. Ctr., Goettingen, Germany, [4]Inst. for Med. Biometry, Informatics and Epidemiology (IMBIE), Faculty of Med., Univ. of Bonn, Bonn, Germany
Disclosure Block: **V. Talevi:** None.
Aging is characterized by a progressive decline of molecular processes which can lead to the development of cardiovascular disease (CVD). MicroRNAs are regulators of gene expression via base-pairing to the target mRNA and they are involved in numerous complex biological processes, including those linked to CVD.
We investigated the association of peripheral blood microRNAs with the main cardiovascular (CV) endophenotypes in a population-based study. In addition, to reveal the biological role of CV-associated microRNAs, we investigated their corresponding targeted genes by microRNA target prediction.
Peripheral blood microRNA expression was acquired through RNA-Seq for 2196 participants from the Rhineland Study, an ongoing population-based cohort study in Bonn, Germany. MicroRNAs expressed in more than 5% of the participants with an average count greater than five were included in the analysis. Blood pressure (BP) and arterial stiffness (total arterial compliance index (TACI) and pulse wave velocity (PWV)) were assessed in all participants. Associations of microRNA expression with CV endophenotypes were determined through linear regression, adjusted for age, sex, batch and traditional cardiovascular risk factors (including BMI, smoking status, diabetes and total cholesterol). False discovery rate (FDR)-adjusted p-values were applied to account for multiple comparisons. We further investigate the microRNA-targeted genes by querying microT-CDS, TargetScan and miRDB databases.
A total of 2174 participants (mean age = 55 years, range 30.0 - 95.0), 56.0% women) were included in the analysis. Out of 666 microRNAs, we found miR-382-5p, miR-654-5p and miR-424-3p significantly (pFDRIn summary, our results suggest that miR-382-5p, miR-654-5p, miR-424-3p are candidate biomarkers for BP, and miR-1292-5p for arterial stiffness trait. Based on microRNA target prediction, we identified a possible role of these microRNAs in CV-related pathways, which must be verified by further microRNA-mRNA coexpression analyses.

PrgmNr 2474 - Characterization of caffeine response regulatory variants in vascular endothelial cells

View session detail

Author Block: **C. Boye**[1], C. Kalita[2], A. S. Findley[1], A. Alazizi[1], R. Pique-Regi[3], F. Luca[3]; [1]Wayne State Univ., Detroit, MI, [2]Dana-Farber Cancer Inst., Boston, MA, [3]Wayne State Univ, Detroit, MI
Disclosure Block: **C. Boye:** None.
The human genome consists mostly of non-coding regions containing regulatory sequences that are important determinants of gene expression. Variants within these regions can contribute to phenotypes by modulating gene expression. These regulatory regions are also key mediators of the response to environmental stimuli. Gene-environment (GxE) interactions may contribute to conditions such as cardiovascular disease (CVD). Caffeine in particular is the most widely consumed stimulant and is known to produce a vascular response. Though results on the role of caffeine in cardiovascular health have been conflicting, previous studies of GxE in gene expression have confirmed the importance of caffeine in coronary artery disease (CAD) risk. In particular eQTLs for artery tissue that colocalize with CAD risk variants are enriched in binding sites for caffeine response factors. To further investigate GxE interactions for caffeine, we treated vascular endothelial cells with caffeine to measure allele-specific effects (ASE) of these regulatory variants using the massively parallel reporter assay Biallelic Targeted STARR-Seq (BiT-STARR-Seq). This approach has the advantage of avoiding confounding trans effects and is not limited by the allele frequency of the variants to be tested. We performed 6 BiT-STARR-seq experimental replicates and investigated 34,420 SNPs computationally predicted to be regulatory variants. In total we found 4,669 variants with significant ASE (FDR

PrgmNr 2475 - Converging evidence for differential regulatory control of *APOEÎµ4* on African versus European haplotypes

View session detail

Author Block: **K. Nuytemans**[1,2], M. Lipkin[1], L. Wang[1,2], D. Van Booven[1], A. J. Griswold[1,2], F. Rajabli[1], K. Celis[1], O. Oron[1], S. Zhang[3], F. Jin[3], M. Argenziano[4], S. F. Grant[5], A. Chesi[6], C. Brown[7], J. Young[1,2], D. M. Dykxhoorn[1,2], M. A. Pericak-Vance[1,2], J. M. Vance[1,2]; [1]John P. Hussman Inst. for Human Genomics, Miami, FL, [2]Dr. John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami, Miami, FL, [3]Case Western Reserve Univ., Cleveland, OH, [4]Univ. of South Florida, Tampa, FL, [5]Children's Hosp. of Philadelphia, Philadelphia, PA, [6]Univ. of Pennsylvania Perelman Sch. of Med., Wallingford, PA, [7]Univ. of Pennsylvania, Perelman Sch. of Med., Philadelphia, PA

Disclosure Block: **K. Nuytemans:** None.

Background: The difference in *APOEÎµ4* risk for Alzheimer disease (AD) between different populations is associated with *APOEÎµ4* local ancestry (LA), with protective effects seen on African versus European or Japanese LA background. Recently, we have shown that *APOEÎµ4* expression in the frontal cortex from European LA AD patients who are homozygous for *APOEÎµ4* is significantly increased relative to AD homozygous *APOEÎµ4* carriers with African LA. Thus, regulatory differences in LA are most likely involved in the protective factor(s) lowering the risk for African carriers of *APOEÎµ4.* We examined LA SNPs with significant frequency differences between African and European/Japanese *APOEÎµ4* haplotypes for areas of differential regulation. Methods: To assess the regulatory potential of enhancer activity in microglial, neuronal and astrocytic cell lines, we performed two different Massively Parallel Reporter Assay (MPRA) approaches assessing haplotype or single variant effects, supplemented with single fragment reporter assays. Additionally, we utilized Capture C analyses in the same cell types to support chromatin interactions with the *APOE* promoter and performed in-silico annotation for the LA region surrounding *APOE*. Results: A region in *TOMM40* spanning introns 2 and 3 was identified to have increased regulatory activity in the European/Japanese LA haplotypes versus the African LA in both astrocytes and microglia. This region overlaps with *APOE* promoter interactions in Capture C in the same cell types. Single variant analyses pinpoints rs2075650, rs157581 and rs59007384 as functionally different between African and European/Japanese *APOE* LA. Conclusion: Both differential regulatory function and Capture C data support an intronic region in *TOMM40* as contributing to the differential *APOE* expression between African and European/Japanese LA, with European/Japanese LA driving higher expression levels. Follow-up functional analyses of the enhancer region is currently ongoing.

PrgmNr 2476 - Genetic Variation in Cis Regulatory Domains and Trans Regulatory Hubs of Immunity

View session detail

Author Block: **D. Avalos**[1], G. Rey[2], A. Ramisch[3], O. Delaneau[4], E. T. Dermitzakis[2]; [1]Dept. of Computational Biology, Univ. of Lausanne, Lausanne, Switzerland, [2]Univ. of Geneva, Geneva, Switzerland, [3]Geneva, Switzerland, [4]Univ. of Lausanne, Lausanne, Switzerland
Disclosure Block: **D. Avalos:** None.
Studying the interplay between genetic variation and regulation of gene expression in immune cells is important to understand how cellular states are modified in various conditions, including immune diseases. Here we built cis and trans maps of regulatory regions with coordinated activity, also named as Cis Regulatory Domains (CRDs) using data from the BLUEPRINT consortium. This dataset includes whole genome sequencing, histone ChIP-Seq, RNA-Seq and DNA methylomics for up to 200 individuals. We were able to discover 7666, 9287, 4947 histone CRDs and 6112, 6053, 5701 methyl CRDs respectively in neutrophils, monocytes and T cells. Only ~35% of CRDs are shared between cell types, indicating that CRDs show a high level of cell-type specificity. Moreover, most CRDs are under strong genetic control: we found 9980 CRD-QTLs in monocytes. CRD-QTL sharing among cell-types varied between 39% and 86%. We discovered 15294 histone and 6185 methyl CRD-gene associations, with only 33% of sharing between cells, which reveals important cell-type specific regulation of gene expression by CRDs. We defined triplets (genetic variant, CRD, and gene) and subsequently found expression-CRD-QTLs: 21528, 17877 and 5457 histone eCRD-QTLs respectively in neutrophils, monocytes and T-cells. We also integrated CRD maps and associations with Promoter Capture Hi-C (PCHi-C) data available for the same cell types to investigate how functional interactions were related to physical 3D proximity. Analyses in trans revealed extensive inter-chromosomal interactions in primary immune cells, with up to 159422 histone CRD associations in neutrophils and 308 histone trans regulatory hubs (TRHs) in monocytes. 2 of these TRHs (in neutrophils and T-cells) were significantly linked to GO terms associated with immune response (64 terms at 5% FDR). We explored the patterns of histone trans CRD association sharing among cell types and found 25% to 58% of sharing.Finally, we integrated the trans networks with QTL data to discover hundreds of trans-eQTLs across cell types. Overlapping our hits with the trans-eQTLs from a meta-analysis in whole blood (eQTLGen Consortium) revealed that from 121 unique trans-eQTLs found in neutrophils, 40 were trans-eQTLs in eQTLGen involving a gene and a variant in LD with our variant. These results support that our data integration strategy is able to discover trans-eQTLs with a good overlap with existing datasets. Together, we show that mapping functional regulatory units using population genomics data allows discovering important mechanisms in the regulation of gene expression in immune cells.

PrgmNr 2477 - Insulin Resistance is associated with *PPARGC1A* promoter DNA-methylation in Ecuadorian women with Turner syndrome

View session detail

Author Block: **F. Alvarez-Nava**, M. Salinas, D. Bastidas, Y. VicuÃ±a, M. Racines-Orbe; Central Univ. of Ecuador, Quito, Ecuador
Disclosure Block: **F. Alvarez-Nava:** None.
We explored peroxisome proliferator-activated receptor-Î³ co-activator 1Î± gene (*PPARGC1A*) promoter DNA methylation status and mitochondrial content in lymphocytes in relation with glucose metabolism in Ecuadorian Women with Turner Syndrome (TS). In a cross-sectional study, a cohort of 34 Ecuadorian patients with TS along with a sex-, age- and body mass index (BMI)-matched reference group was studied. Insulin resistance and secretion indices were calculated. After bisulphite treatment of peripheral blood lymphocyte genomic DNA, a real-time methylation-specific PCR was used to determine the *PPARGC1A* methylated DNA/unmethylated DNA ratio in selected CpGs. Remarkably, *PPARGC1A* methylated DNA/unmethylated DNA ratios in the study group were significantly higher, compared to the reference group (1.036 vs 0.6097; 95% CI: 0.14 to 0.69; *P*=0.001). This ratio was directly correlated with overweight/obesity and inversely correlated with the biochemical features of insulin resistance (HOMA-IR and Matsuda indices). The mean mtDNA copy numbers of the study and reference groups were estimated as 27.15 Â± 3.21 and 49.93 Â± 6.699, respectively, and was indirectly correlated with methylated DNA/unmethylated DNA ratio of the *PPARGC1A* promoter (r= -0.78; 95 CI -0.9 to -0.55; *P*=PPARGC1A, independently of age or BMI, and contributed to 20% of the total variability in HOMA-IR index in TS subjects. Our findings suggest that PPARGC1A promoter DNA methylation status and lower mitochondrial number affect the metabolic phenotype in TS subjects.

PrgmNr 2478 - Investigating the role of a cardiac enhancer in the development of heart failure

View session detail

Author Block: **M. Htet**, M. Arvanitis, B. Lin, H. Gangrade, E. Tampakakis; Johns Hopkins Univ. Sch. of Med., Baltimore, MD
Disclosure Block: **M. Htet:** None.
Introduction: Alpha actinin-2 (ACTN2) is a major cytoskeletal protein that plays a critical role in maintaining the structural and functional integrity of the sarcomere. *ACTN2* mutations, although rare, have been shown to be associated with various types of cardiomyopathy. Using genome-wide association and multi-omic approaches, we recently identified non-coding variants that showed a strong association with heart failure (HF). Two of the variants are within an evolutionary conserved transcriptional enhancer region that regulates the *ACTN2* gene in human stem cell derived cardiomyocytes (hPSC-CMs). However, the role of cardiac enhancers in the development of HF remains unknown, therefore we used engineered heart tissues (EHT) from hPSC-CMs to investigate the effects of the ACTN2 enhancer. **Methods:** We used hPSC-CMs and performed morphologic and functional analyses by immunostaining, calcium transient measurements and quantitative assays such as western blotting, proteomics, qPCR and single cell transcriptomics. Next, we utilized CRISPR interference (CRISPRi) and chromatin immunoprecipitation (ChIP) to analyze the transcriptional regulation of the *ACTN2* enhancer. Finally, we measured force generation using EHTs. **Results:** We first engineered hPSC-CMs carrying an *ACTN2* enhancer deletion. This resulted in decreased *ACTN2* gene and protein expression, and cardiomyocytes developed myofibrillar disarray, hypertrophy, lower beating rates and suppressed calcium transients. Moreover, EHTs demonstrated reduced mechanical force. Using CRISPRi and ChIP we found that the transcription factor MEF2c binds a short DNA sequence containing an enhancer variant to regulate *ACTN2* expression. Single cell RNA-Seq analysis of hPSC-CMs treated with isoproterenol to model the hyperadrenergic state observed in HF, revealed the induction of pathways involved in protein quality control, actin fragmentation and apoptosis. We have also recently developed a mouse model to address the effect of the enhancer *in vivo*. **Conclusions:** Our study confirms that a conserved and clinically relevant enhancer region can effectively regulate *ACTN2*, and variants within that region can have detrimental consequences on CMs which can contribute to HF.

PrgmNr 2479 - Linking GWAS signals to target genes in skeletal muscle using single-nuclei multi-omic analyses

View session detail

Author Block: **C. Ventresca**[1], P. Orchard[2], N. Manickam[3], K. Gallagher[4,5], S. Parker[2]; [1]Dept. of Human Genetics, Univ. of Michigan, Ann Arbor, MI, [2]Univ. of Michigan, Ann Arbor, MI, [3]Dept. of Computational Med. and Bioinformatics, Univ. of Michigan, Ann Arbor, MI, [4]Dept. of Surgery, Univ. of Michigan, Ann Arbor, MI, [5]Dept. of Microbiol. and Immunology, Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **C. Ventresca:** None.
Type 2 diabetes (T2D) is a complex disease that arises from a combination of genes and environmental factors. Previous genome wide association studies (GWAS) of T2D have identified >600 independent genetic signals, but most fall within non-coding regions of the genome, which makes identifying underlying mechanisms difficult. These signals can be cell-type specific, which adds another layer of complexity to nominating mechanisms. Previous studies have established that chromatin architecture can pre-determine transcriptional responses and is therefore a critical molecular property of cell state potential. With this in mind, to investigate gene expression and chromatin accessibility in a more integrated manner, we generated single-nucleus multi-omic (joint RNA and ATAC profiling on the same nucleus) profiles in skeletal muscle. From this experiment we generated 7,613 high-quality pass-QC nuclei with a mean of >2k RNA UMIs and >48k ATAC reads per nucleus. From this sample we have identified nine different cell types ranging in abundance from 46% (type 1 muscle fibers) to 0.08% (smooth muscle) of all nuclei. This data is being used to link chromatin peaks to target genes with higher precision, since both RNA and ATAC datasets originate from the same nucleus. By performing peak-gene and peak-peak association analyses, we are linking GWAS signals to effector transcripts. We aim to use this multi-omics approach to compare the genetic control of gene expression and chromatin accessibility with the goal of nominating cell-specific skeletal muscle mechanisms at T2D and related trait GWAS signals.

PrgmNr 2480 - New epigenomic panel to detect mendelian disorders in Brazilian health system

View session detail

Author Block: **G. S. Carvalho**[1], L. L. Vieira[1], Y. G. Gasparini[2], V. Almeida[3], A. M. Nascimento[1], B. M. Wolff[1], M. Silva[2], M. M. Montenegro[4], L. D. Kulikowski[5]; [1]Faculdade de Med. da Univ.e de SÃ£o Paulo, SÃ£o Paulo, Brazil, [2]Univ. of SÃ£o Paulo, SÃ£o Paulo, Brazil, [3]Faculdade de Med. da Univ.e de SÃ£o Paulo, Sao Paulo, SP, Brazil, [4]FMUSP, Sao Paulo, Brazil, [5]Univ. de Sao Paulo, Sao Paulo, Brazil
Disclosure Block: **G.S. Carvalho:** None.

Methylation is a main epigenomic tool for gene expression control in mammals. Through this mechanism, the genome presents a variable expression pattern between tissues, cells and alleles from different parental origin. Thus, structural variations in DNA, such as CNVs and point mutations, are not the only or definitive causes for clinical phenotypes. Several imprinting disorders, as well as recurrent changes in the methylation pattern in multifactorial diseases, are reported in the literature. However, diagnosis by epigenetic investigation tools still does not have a broad and consolidated pipeline, consequently making it difficult to use this in diagnostic routine. In this study, we sought to identify and validate, through a meta-analysis, the existence of epi-signatures for 16 different Mendelian diseases related to genomic imprinting or epigenomic changes in methylation. The data were obtained by searching public databases, such as Gene Expression Omnibus (GEO), and in several published studies. It was possible to identify the existence of epi-signatures for 11 Mendelian disorders. All diseases related to genomic imprinting (Angelman, Prader-Willi, Beckwith-Wiedemann, Silver-Russel, Kagami-Ogata and Temple syndromes) has significant data for correlation with characteristic DMRs. Kabuki, CHARGE, Claes-Jensen, Fragile X and Williams-Beuren syndromes do not have an etiology related to genomic imprinting, although present DMRs in recurrent regions. With these results, it was possible to develop an epigenomic array panel for screening epigenomics changes related to these diseases. Even among multifactorial diseases, there is an epigenetic profile. Thus, the use of epigenomic array screening platforms, can be considered promising for a new level of diagnostic analysis, providing a multidimensional characterization of the genomic structure. Furthermore, the consolidation of epigenomic diagnostic methodologies into a single tool, such as the one proposed in this study, provides economic, laboratory and diagnostic advantages.

PrgmNr 2481 - Promoter methylation analysis in candidate genes of severe preterm birth

View session detail

Author Block: **S. Pereyra**[1], A. Sardina[1], R. Neumann[2], C. May[2], B. Bertoni[1], R. Sapiro[3], M. Cappetta[1]; [1]Departamento de GenÃ©tica, Facultad de Med.. Univ. de la RepÃºblica, Montevideo, Uruguay, [2]Dept. of Genetics, Coll. of Med., Univ. of Leicester, UK, Leicester, United Kingdom, [3]Departamento de HistologÃa y EmbriologÃa, Facultad de Med., Univ. de la RepÃºblica, Montevideo, Uruguay
Disclosure Block: **S. Pereyra:** None.
Preterm birth (PTB), defined as the delivery of an infant before 37 weeks of gestation, results from the interaction of genetic and environmental components and constitutes a complex multifactorial syndrome. The etiology of PTB is not completely understood because of the complex interplay of genetic, environmental, and host factors. Therefore, it is key to develop approaches that integrate genetic, expression, epigenetic and epidemiological data, to identify biomarkers that allow progress in translational medicine.To this end, we developed an analysis pipeline to study the methylation status of gene promoters. In this work, we tested this workflow in candidate genes for severe premature labor, selected because they were previously found differentially expressed in chorioamniotic tissue in preterm birth in an Uruguayan population. To test the methylation levels in preterm birth gene promoters, we analyzed chorioamniotic tissue from 6 severe preterm birth and 4 controls delivered at term. Genomic DNA was treated with sodium bisulfite and 13 gene promoters of preterm birth candidate genes were amplified: *ANKRD2*, *ACCS*, *BIRC3*, *WNT1*, *CXCL2*, *NRN1*, *STEAP1*, *EGR3*, *MAMDC2*, *CNTNAP3B*, *GK*, *PIK3AP1* and *MIR155HG*. Amplicons, with an average length of 1000bp, were sequenced using the MinION platform (Oxford Nanopore). We pooled amplicons of each individual, prepared libraries using PCR Barcoding kit and sequenced them in Spot-ON 106D R9 flow cell. Basecalling and debarcoing were carried out in MinKnow, reads were filtered by quality with Nanofilt. We used Bismark with custom score parameters to align bisulfite treated sequenced reads to a human reference genome (hg38) and to extract methylation status. CpGs were filtered by coverage (30x) and analyzed in R software. We detected hypomethylated regions in severe preterm birth as compared to controls. We found differential methylation in promoters between patients and controls in genes *NRN1* and *MAMDC2* (Wilcoxon test, p

PrgmNr 2482 - Using cellular deconvolution to investigate cell subtype proportions in cortical gene expression data in schizophrenia

View session detail

Author Block: **R. Mahoney**[1], C. Seoighe[2], D. Morris[1]; [1]Ctr. for Neuroimaging, Cognition and Genomics, Discipline of Biochemistry, Natl. Univ. of Ireland Galway, Galway, Ireland, [2]Sch. of Mathematics, Statistics and Applied Mathematics, Natl. Univ. of Ireland Galway, Galway, Ireland
Disclosure Block: **R. Mahoney:** None.
Schizophrenia is a psychiatric disorder that affects 1% of adults and is a major global health problem in that only 13.5% of affected individuals achieve full recovery criteria. Altered gene expression in the brain has previously been associated with neuropsychiatric disorders but this research has primarily focused on bulk tissue. Further analysis of data from bulk tissue to isolate and study predicted cell types based on gene expression profiles could uncover further insights into the altered expression of genes and pathways that contribute to schizophrenia etiology. Cell subtype deconvolution is used to estimate the proportion of cell subtypes present in bulk expression data. Here, we reanalyzed gene expression data from the PsychENCODE consortium (n= 558 schizophrenia cases and 1039 controls; cortical samples) by using new single cell sequencing data that allowed for an improved cell subtype deconvolution analysis.
New single cell data for ~30,000 cells was added to the original dataset to give an increased sample of ~62,000 cells. In comparison to the original analysis, we observed alterations in astrocyte proportions between cases and controls, particularly for two distinct astrocyte populations, only one of which had significantly different proportions between cases and controls (p=

PrgmNr 2483 - Cas9 targeted enrichment of mobile elements using nanopore sequencing

View session detail

Author Block: **W. Zhou**[1], T. McDonald[1], C. Castro[1], C. Mumm[1], J. Switzenberg[1], R. Mills[2], A. P. Boyle[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ Michigan Med Sch, Ann Arbor, MI
Disclosure Block: **W. Zhou:** None.
Mobile element insertions (MEIs) are repetitive genomic sequences that contribute to genetic variation as well as various genetic disorders. Targeted and whole-genome approaches using short-read sequencing have been developed to identify reference and non-reference MEIs; however, the read length hampers the detection of these elements in complex genomic regions. Here, we pair Cas9-targeted nanopore sequencing with customized computational methodology (Nano-Pal) to capture active MEIs in human genomes. We design the guide RNAs for repetitive MEIs using unique subfamily-specific sequences. We demonstrate parallel enrichment for distinct classes of MEIs, averaging 44% of reads on-targeted signals (64% L1Hs, 52% *Alu*Yb, 64% *Alu*Ya, 10% SVA_F, 3% SVA_E, and 56% pooled) and exhibiting a 13.4-54x enrichment over whole-genome approaches. We show an individual MinION flow cell can recover most MEIs (97% L1Hs, 93% *Alu*Yb, 51% *Alu*Ya, 99% SVA_F, and 65% SVA_E). In a well-characterized benchmark genome, NA12878, we identify seventeen non-reference MEIs with retrotransposition hallmarks overlooked by modern, long-read analysis pipelines, primarily in repetitive genomic regions. We further provide the CpG methylation profiles of reference and non-reference MEIs derived from individual nanopore reads. This work introduces the utility of nanopore sequencing for MEI enrichment and lays the foundation for rapid discovery and potential functional analysis of elusive, repetitive genetic elements.

PrgmNr 2484 - Comparative transcriptomics of the human lung and epididymis epithelia

View session detail

Author Block: **A. Paranjapye**, S. Nandikolmath, J. Kerschner, S. Yin, S-H. Leir, A. Harris; Case Western Reserve Univ., Cleveland, OH
Disclosure Block: **A. Paranjapye:** None.
The different cell lineages within epithelial tissues have distinct gene expression profiles. These unique transcriptomes confer the identity and region-specific diversity in epithelial function, for example in human airway/bronchial and male genital duct/epididymal epithelium (HBE and HEE). Here, we perform comparative analysis of whole epithelium RNA-seq and of single cell (sc)RNA-seq to identify patterns of gene expression shared and contrasted between these two similar tissues. Gene ontology and gene set enrichment analysis of whole epithelium datasets identified genes and molecular processes specific for the lung or epididymis, which may help explain mechanisms behind their unique biological functions. We next performed the same analysis between clusters in HBE and HEE single cell data to generate similarity matrices for gene expression. Varying degrees of correspondence in the inter-cell comparison offers insights into the differentiation of known cell types between the lung and epididymis. Integration of the datasets further reveals the concordance and discordance of gene expression profiles between stratified clusters. Finally, we annotated and performed DNA footprinting of HBE and HEE open chromatin data to elucidate similarities at the transcription factor binding level.

PrgmNr 2485 - Detecting differentially methylated regions in individual methylomes

View session detail

Author Block: **C. Hansen**, M. Janecka; Mount Sinai, New York, NY
Disclosure Block: **C. Hansen:** None.
Recent exome sequencing studies have demonstrated a number of highly penetrant rare and *de novo* variants and copy number variations underlining rare monogenic forms of autism spectrum disorder (ASD) - suggesting a multiple rare variants etiology. Rare and extreme deviations in individual methylomes, called epivariations, have also been linked with the risk of neurodevelopmental disorders, including ASD. Epivariations may be a consequence of rare or *de novo* genetic events, but also of environmental/stochastic effects in cases with no discernable genetic cause - explaining an otherwise unknown etiology. An epivariation is defined as an extreme methylomic deviation across a genomic region, compared to a set of reference methylomes. In this study we devised a novel approach to detect and quantify epivariations from differentially methylated regions (DMRs) in individual methylomes assessed by DNA methylation array. Our method uses individual genome-wide M-values, which we z-score transform by the cohort median and median absolute deviation. This ensures similar variance across all CpG sites, and is robust towards outliers. We combine all possible sequential combinations of z-scores in a genomic region with the Stouffer-Liptak method, with the Kechris adjustment for spatial auto-correlation between neighbouring CpG sites. The sequences of non-overlapping CpG sites resulting in the highest absolute combined z-scores indicate the optimal boundaries of each DMR. Within these boundaries we assess methylation deviation across the DMR by taking the mean of all z-scores. This method robustly detects DMRs across samples and quantifies their deviance unbiased by the number of CpG sites included in the DMR - which may vary between array designs and genomic regions. We use this information to bin DMRs by their deviance and evaluate their frequency and functional properties in relation to neurodevelopmental phenotypes. Individual DMR pathogenicity is evaluated from a list of genes previously linked with psychiatric disorders, methylomic developmental sites as well as loss of function intolerance scores. We find that epivariations underlie phenotypical heterogeneity in ASD, mark monozygotic twins discordant for disease, and that moderate DMRs bridge the gap between the small but common methylomic deviations observed in epigenome-wide association studies and the rare high impact epivariations observed in previous studies.

PrgmNr 2486 - Direct assessment of impact of transcription factors on enhancer activity

View session detail

Author Block: **M. Das**[1], A. Hossain[1], M. C. Jensen[2], V. Pounraja[3], S. Mogre[1], A. Tyryshkina[3], J. Mao[1], S. Girirajan[4]; [1]The Pennsylvania State Univ., University Park, PA, [2]Pennsylvania State Univ, University Park, PA, [3]The Pennsylvania State Univ., State College, PA, [4]Pennsylvania State Univ., University Park, PA

Disclosure Block: **M. Das:** None.

Identification and quantification of cis-regulatory elements have provided new insights into their roles in complex disease. There has been significant progress in the ability to assess enhancer function, enhancer-gene interactions, and effects of enhancer variants on phenotypes using both experimental and computational approaches. However, direct functional evidence of enhancer activity in response to transcription factor binding is limited. Traditionally, enhancers have been identified using ChIP-seq of enhancer-binding transcription factors or histone marks as well as profiling open-chromatin regions using DNase-seq and ATAC-seq. Further, massively-parallel reporter assays including STARR-seq have been used to directly quantify enhancers by transfecting candidate enhancer libraries into a host cell and sequencing self-transcribed enhancer fragments. Here, using a series of STARR-seq assays on HEK293T cell lines deleted for a suite of six transcription factors compared to control lines, we provide direct evidence of association of transcription factors towards enhancer activity. We hypothesized that deletion of each transcription factor will alter activity of specific enhancers and will help categorize enhancers that are unique to or shared across multiple transcription factors. We generated a comprehensive library of 46,000 enhancer fragments curated from both cell-type specific DNase-seq and ChIP-seq sites as well as global enhancer marks for the six transcription factors. Using CRISPR-Cas9, we then performed genomic deletions to obtain significant (>75%) depletion of the transcription factor proteins, as assessed using western blots. We will use STARR-seq to assess and quantify enhancer activity for each deletion line, grouped by their roles in gene regulation. For instance, we will test enhancer activity changes in deletion lines for canonical transcription factors such as ATF2, FOXA1 and CTCF. Similarly, we will assess enhancer activity for lines individually deleted for Wnt signaling genes LEF1 and TCF7L2. We will finally assess enhancer activity after deleting neurodevelopmental disease associated SCRT1 (a neurodevelopmental gene) and the 16p12.1 deletion. We will use enhancer activity change in combination with differential gene expression data to uncover the genetic etiology of complex neurodevelopmental disease. Comparison of the activity patterns across groups will provide a framework for understanding enhancer activity signatures demonstrated by distinct gene regulatory groups. Our strategy will help formulate and refine enhancer-based gene regulatory networks responsible for disease complexity.

PrgmNr 2488 - Escape from X-inactivation in twins exhibits intra- and inter-individual variability across tissues

View session detail

Author Block: **A. Zito**[1,2], A. Visconti[1], N. Rossi[1], A. L. Roberts[1], R. Andres-Ejarque[3], S. Nardone[4], J. E. S. Moustafa[1], M. Falchi[1], K. S. Small[1]; [1]Twin Res. & Genetic Epidemiology, King's Coll. London, London, United Kingdom, [2]Present address: Massachusetts Gen. Hosp., Boston, MA, [3]St Johnâ□□s Inst. of Dermatology, Kingâ□□s Coll. London, London, United Kingdom, [4]Dept. of Med., Beth Israel Deaconess Med. Ctr., Harvard Med. Sch., Boston, MA

Disclosure Block: **A. Zito:** None.

X-chromosome inactivation (XCI) silences one X in females to balance the unequal X-linked transcriptional dosage between the sexes. However, over 15% of X-genes escape XCI and is biallelically expressed. Our knowledge of XCI escape in humans largely rely on sex-differences, hybrid cell lines, and epigenetic marks, which are all indirect proxies of escape. Differently, tissue samples with skewed XCI enable the detection and measurement of escape directly in females. At present, the incidence and variability of escape across cells, tissues and individuals are not well known. The extent to which genetics and environment influence escape is also largely unknown. Using RNAseq and DNAseq data in a multi-tissue dataset sampled from 248 skewed female twins, we investigated escape prevalence and variability across fat and skin tissues, lymphoblastoid cell lines (LCLs), and purified immune cells (monocytes, B, T-CD4$^+$, T-CD8$^+$, NK), and individuals.

Solid tissues exhibit up to 13% higher incidence of escape than LCLs. 159 genes escaped XCI in least one tissue. This set includes 54 novel candidate escapees, of which 35% are long-non-coding RNAs. 24 genes constitutively escaped XCI across tissues, while a separate set of 51 genes exhibited tissue-restricted escape. Within an individual, there are both genes escaping XCI in all tissues, and genes showing tissue-restricted escape. Analysis of inter-individual variability revealed 40 genes (e.g. *BTK*, *CD99L2*) exhibiting consistent escape levels across females in at least one tissue, and 62 genes (e.g. *DDX3X*, *KDM6A*, *UBA1*) exhibiting inter-female variability in multiple tissues. Escape is heterogeneous across immune cell types, with higher incidence in lymphocytes than monocytes. There are both genes escaping XCI in multiple immune cell types, and immune cell type-specific escapees. Protein-protein network analysis highlighted that escapees interact with other proteins on a genome-wide scale, and are involved in varied processes and pathways such as epigenetic control of gene activity. Monozygotic (MZ) co-twins share significantly more similar escape levels than dizygotic (DZ) co-twins (corr.MZ=0.6; corr.DZ=0.46; P

PrgmNr 2489 - Genetic variants regulating microRNA expression in the developing human neocortex

View session detail

Author Block: **M. Lafferty**[1], N. Aygun[2], D. Liang[1], J. Wolter[1], D. H. Geschwind[3], L. de la Torre-Ubieta[4], J. Stein[2]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]UNC Chapel Hill, Chapel Hill, NC, [3]Univ California los Angeles, Los Angeles, CA, [4]Univ. of California Los Angeles, Los Angeles, CA
Disclosure Block: **M. Lafferty:** None.
Genome-wide association studies (GWAS) have revealed a number of genomic loci associated with risk for neuropsychiatric disorders and brain-related traits, but for many of these loci, the causal mechanisms are unknown. Using expression quantitative trait loci (eQTL) data, many of the risk loci can be linked to protein-coding genes or lincRNAs. However, the mechanism underlying a significant proportion of loci remain unexplained. MicroRNAs (miRNAs) are poorly measured in standard eQTL studies yet have important influences on neurogenesis and have been found to be differentially expressed in brain tissue from patients with neuropsychiatric disorders. By linking these miRNA-eQTLs to previously defined mRNA-eQTLs and genome-wide significant loci for brain traits, we aim to reveal causal mechanisms by which common genetic variation influences gene regulation and risk for brain-related traits.

Here, in 212 donors, we identified 85 miRNA-eQTLs associated with expression of 70 miRNAs (emiRs) in human cortical tissue during mid-gestation. Moreover, 13 miRNA-eQTLs were associated with expression of previously unannotated miRNAs which we discovered here in developing cortical tissue. We found enrichment of miRNA-eQTL signal within active transcription start sights and among chromatin associated with transcription. Co-localization with mRNA-eQTLs within the same population of tissue samples yielded 23 miRNA-eQTLs that co-localize with 33 mRNA-eQTLs. Of the 23 miRNA-eQTL co-localizations, 18 of the emiRs are within the expressed host-gene of the mRNA-eQTL, indicating a co-transcriptional mechanism by which a mRNA-eQTL also influences miRNA expression. Co-localization of miRNA-eQTLs with GWAS summary statistics yielded one robust co-localization of miR-4707-3p expression to educational attainment and head size phenotypes. MiR-4707-3p also co-localizes with a mRNA-eQTL of HAUS4, which is the host-gene of miR-4707 (located within exon 1 of HAUS4). Expression of HAUS4, a protein-coding gene known to play a role in cell-cycle regulation and proliferation, is anti-correlated with miR-4707-3p expression, consistent with miRNA biogenesis from an exon of the host-gene.

Ongoing work will identify miR-4707-3p targets and the effect of miR-4707-3p and HAUS4 overexpression on proliferation in human neural progenitor cell lines. Using local-miRNA-eQTLs, we propose novel mechanisms by which genetic variation influences brain related traits.

PrgmNr 2490 - Mapping chromatin interactions within the *CYP3A* gene locus

View session detail

Author Block: **J. M. Collins**[1], D. Wang[2]; [1]Pharmacotherapy and translational research, Univeristy of Florida, Gainesville, FL, [2]Univ. of Florida, Gainesville, FL
Disclosure Block: **J.M. Collins:** None.
The cytochrome P450 CYP3A4 is the most abundant drug-metabolizing enzyme in the liver, metabolizing nearly half of the commonly prescribed medications. CYP3A4 expression is highly heritable and shows significant variation between individuals, but to date, functionally classified SNPs do not explain most of the variation, and cis-acting regulatory elements controlling CYP3A4 transcription remain uncertain. CYP3A4 is located within a cluster of four CYP3A genes (CYP3A4, 3A5, 3A7 and 3A43) that have different tissue and developmental expression profiles. Previously, using chromatin conformation capture (4C) assays and CRISPR-mediated deletion, we identified and characterized several cis-acting regulatory elements controlling CYP3A4 expression in the human liver. To further elucidate potential distal enhancers/regulatory elements and their interactions that regulate expression of the four CYP3A genes, we conducted 4C experiments targeting all four CYP3A promoters as viewpoints. We prepared 4C templates from several cell types that show different CYP3A expression patterns: LS174T (intestinal epithelial carcinoma), Huh7 (hepatoma), and human primary hepatocytes from both black and white donors. Moreover, we performed ATAC-Seq on the same hepatocytes from black and white donors to explore accessible chromatin indicative of regulatory regions. By leveraging 4C, ATACseq, and publicly-available ChIP-Seq datasets for enhancer marks: the histone modifications H3K4me1 and H3K27ac and the p300 histone acetyltransferase protein, we have identified several putative regulatory regions across the CYP3A locus. We are in the process of validating these peaks and expect to draw a chromatin interaction map for the CYP3A locus. The identification of cis-acting regulatory elements controlling CYP3A gene expression is a critical step toward understanding genetic and epigenetic factors contributing to large inter-person variabilities in CYP3A gene expression, with the promise of identifying biomarkers for personalized drug therapy

PrgmNr 2491 - Muscle single nuclei association of cell type proportions, cell type chromatin accessibility and gene expression with sex and T2D-related physiological traits

View session detail

Author Block: **S. Hanks**[1], D. Ciotlos[1], A. Varshney[1], M. Erdos[2], N. Manickam[1], A. U. Jackson[3], J. Okamoto[4], H. M. Stringham[1], P. Orchard[3], N. Narisu[2], L. Bonnycastle[2], M. Laakso[5], J. Tuomilehto[6], T. A. Lakka[5], K. L. Mohlke[7], M. Boehnke[3], H. A. Koistinen[6], F. S. Collins[2], S. Parker[3], L. J. Scott[8]; [1]Univ MICHIGAN, Ann Arbor, MI, [2]NHGRI (NIH), Bethesda, MD, [3]Univ. of Michigan, Ann Arbor, MI, [4]Madison, CT, [5]Univ of Eastern Finland, Kuopio, Finland, [6]Finnish Inst for Hlth.and Welfare, Helsinki, Finland, [7]Univ North Carolina, Chapel Hill, NC, [8]Univ Michigan, Ann Arbor, MI

Disclosure Block: **S. Hanks:** None.

Skeletal muscle is quantitatively the largest organ to regulate metabolism. Our goal is to understand the interplay of skeletal muscle tissue cell type composition and cell state with sex, physical activity and T2D-related physiological traits. We profiled single nucleus resolution chromatin accessibility (snATAC-seq) and gene expression (snRNA-seq) across 287 Finnish individuals with frozen human vastus lateralis skeletal muscle biopsies. We clustered 224,690 snATAC-seq and 190,291 snRNA-seq nuclei using LIGER and identified three main muscle cell clusters corresponding to Type 1 (slow twitch oxidative), Type 2A (fast twitch oxidative), and Type 2X (fast twitch glycolytic) muscle fibers, as well as 10 additional cell types. In each cell type, we used a negative binomial regression model to test for association of the number of nuclei (RNA and ATAC-seq) with sex, age, physical activity levels and T2D-related traits (accounting for total nuclei number). Using linear regression, we tested separately for association of snATAC-seq peaks levels and gene expression with the same set of traits. We used FDR -7), higher energy expended was associated with higher proportion of smooth muscle cells (p= 7.6×10^{-6}), and higher 24 hr physical activity was associated with higher proportion of endothelial cell (p= 8.0×10^{-4}). For the most abundant cell types (in order of overall abundance Type 1, Type 2A and Type 2X), being male (p= 8.7×10^{-13}) and having a higher fasting insulin concentration (p= 3.2×10^{-8}), were associated with higher proportion of Type 2X muscle fiber nuclei, whereas female sex (p=.00022) and lower fasting insulin (p=7.6×10^{-4}) were associated with higher proportion of Type 1 muscle fiber nuclei. No physiological traits were associated with Type 2A fiber nuclei proportions. We tested for association of chromatin accessibility and gene expression with physiological traits in each cell type. Type 2A muscle fiber nuclei had between 1.2->10 fold more ATAC-Seq peaks and genes were associated with sex and fasting insulin than Type 1 and 2X muscle fiber nuclei. This suggests that cell types with relatively low variability in proportion of nuclei by phenotypic trait, such as Type 2A muscle fibers, can have relatively high variability in gene expression and chromatin accessibility for the same traits. Deeper understanding of the interplay between physiological traits, cell type composition and cell states will provide clues to metabolic disease progression.

PrgmNr 2492 - Quantifying the shared genetic effects on the regulation of expression and protein levels in related individuals

View session detail

Author Block: **T. Dupuis**[1], J. M. Soria[2], J. C. Souto[3], A. Martinez[4], J. M. Schwenk[5], E. T. Dermitzakis[6], E. R. Pearson[1], A. ViÃ±uela[7], A. Brown[8]; [1]Univ. of Dundee, Sch. of Med., Dundee, United Kingdom, [2]Hosp. de Sant Pau, Barcelona, Barcelona, Spain, [3]Hosp Sant Pau, Barcelona, Barcelona, Spain, [4]H. de la Santa Creu i Sant Pau, Barcelona, Barcelona, Spain, [5]Dept. of Protein Sci., KTH - Royal Inst. of Technology, Solna, Sweden, [6]Univ. of Geneva, Geneva, Switzerland, [7]Inst. of Genetic Med., Intl. Ctr. for Life, Newcastle Univ., Newcastle upon Tyne, United Kingdom, [8]Univ. of Dundee, Dundee, United Kingdom

Disclosure Block: **T. Dupuis:** None.

RNA-seq experiments can be used to identify the genes which mediate GWAS genetic effects on disease risk. However, it will not explain those variants that act on regulatory processes downstream of transcription. Studies have found low correlations between expression and protein levels, but these could be due to environmental and technical factors. Here, we evaluate the effect of eQTLs on translation by quantifying shared genetics regulation of expression and protein levels. We estimated the heritability and genetic correlations (rho_G) of whole-blood gene expression (RNA-seq) and plasma proteins (Olink) in a pedigree from the GAIT2 project (N=67, 90 proteins, expression of 16748 genes). We identified 42 expression-protein pairs with high reliable estimates of rho_G. We observed a low phenotypic correlation (median $|rho_P|$ of 0.11) but high absolute genetic correlations median $|rho_G|$=0.25, Wilcoxon test $|rho_G| > |rho_P|$, p=2.5e-6). This implies that for many genes there is a common set of variants regulating both expression and protein levels. One example is *RETN*, with rho_G=0.78 and rho_P=0.49. The gene has been previously associated with LDL cholesterol levels, and it is known to have a variant (rs62109837) associated with both expression and protein levels. Overall, for a considerable proportion of genes (18/42) the genetic correlation between expression and protein levels is negative, implying genetic effects of opposite directions. Of known genetic effects on expression and protein levels, a far smaller proportion have opposite signs. For instance, studies in DIRECT[1] have identified 48 variants with effects on both the expression and protein levels, among which only 4 had opposite directions of effect, meaning more work is needed to understand the mechanism behind this finding. Our results indicate that the low phenotypic correlation between expression and protein levels is mainly driven by environmental and technical factors, and that there is a substantial sharing of genetic effects between expression and proteins abundance.

1: Vinì□uela, A. (2021). Genetic analysis of blood molecular phenotypes reveals regulatory networks affecting complex traits: A DIRECT study. MedRxiv (https://doi.org/10.1101/2021.03.26.21254347)

PrgmNr 2493 - Sex-specific and Generational Effects of Alcohol Use and Tobacco Smoking on Epigenetic Age Acceleration in the Michigan Longitudinal Study

View session detail

Author Block: **W. Zhao**[1], A. Carter[1], C. Bares[2], L. Lin[1], B. G. Reed[2,3], M. Bowden[2,4], R. A. Zucker[5,6], J. A. Smith[1,7], J. B. Becker[5,8]; [1]Dept. of Epidemiology, Sch. of Publ. Hlth., Univ. of Michigan, Ann Arbor, MI, [2]Sch. of Social Work, Univ. of Michigan, Ann Arbor, MI, [3]Dept. of Women's and Gender Studies, Univ. of Michigan, Ann Arbor, MI, [4]Michigan Med., Univ. of Michigan, Ann Arbor, MI, [5]Dept. of Psychology, Univ. of Michigan, Ann Arbor, MI, [6]Dept. Psychiatry and Addiction Ctr., Univ. of Michigan, Ann Arbor, MI, [7]Survey Res. Ctr., Inst. for Social Res., Univ. of Michigan, Ann Arbor, MI, [8]Michigan NeuroSci. Inst., Univ. of Michigan, Ann Arbor, MI

Disclosure Block: **W. Zhao:** None.

Background: Excessive alcohol use and tobacco smoking are risk factors for poor health in both men and women, but use patterns and relationships with diseases and mortality differ between sexes. The impact of these lifestyle factors on biological markers, such as DNA methylation, may also be different by sex. Whether parental alcohol or tobacco intake has effects on the epigenetic profiles of children that persist into adulthood is not well studied. The epigenetic age (DNAmAge), an aggregated measure over many aging-related methylation sites across the genome, is an emerging biomarker of aging. The difference between DNAmAge and chronological age, herein referred to as epigenetic age acceleration (DNAmAA), correlates with a variety of health outcomes. In this study, we assessed the sex-specific effects of alcohol and tobacco use, as well as parental alcohol/tobacco use, on four measures of DNAmAA in a longitudinal cohort with multiple generations.

Methods: In 42 parents and 45 offspring from the Michigan Longitudinal Study (MLS), four measures of DNAmAA (HannumAA, HorvathAA, PhenoAA and GrimAA) were estimated from saliva samples. Linear mixed models were used to assess the associations between alcohol use or tobacco smoking and DNAmAA in males and females separately. Alcohol use/tobacco smoking by sex interaction models were used to assess whether the effects differ by sex. In the offspring only, the association between parental alcohol use and tobacco smoking (when the offspring was 12 years old or younger) with adult DNAmAA was also examined.

Results: In females, but not males, each additional alcoholic drink per month was associated with a 0.047-year increase in GrimAA (p = 0.046). Males who were current tobacco smokers had an increased HannumAA of 2.19 years (p = 0.043); this was not found in females. Heavy paternal alcohol use was associated with a 3.75-year increase in GrimAA (p = 0.046) among offspring. Parental tobacco smoking was not associated with any DNAmAA measure.

Conclusions: This study found sex-specific effects of alcohol use and tobacco smoking on epigenetic age acceleration. Heavy paternal alcohol use when the offspring were 12 years old or younger was associated with increased offspring age acceleration in adulthood. These findings suggest that paternal alcohol use influences adult DNAmAA among offspring, and that sex-specific differences in the effects of alcohol use and tobacco smoking on DNAmAA should be further examined.

PrgmNr 2494 - Stressing out epigenetics: How chronic stress alters the epigenetic landscape during neurodevelopment

View session detail

Author Block: **J. N. Kuehner**; Emory Universtiy, Atlanta, GA
Disclosure Block: **J.N. Kuehner:** None.
Stress is the multi-level response an organism has to an environmental stimulus, and when experienced chronically, has negative consequences such as the development of anxiety or depressive behaviors. It is well documented that epigenetics has critical roles in brain development and function; however, a major gap in our knowledge about chronic stress and depression are what rolls stress-induced epigenetic alterations could be contributing to. Whether DNA modifications and their machinery are involved in stress response has yet to be elucidated. 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) are two of the most abundant DNA modifications in the mammalian genome and have been implicated in stress response. However, a genome-wide analysis of the 5mC or the 5hmC landscape in response to chronic stress have yet to be rigorously investigated. To determine how chronic stress can alter the epigenetic profile in various brain regions during mouse brain maturation, we employed the chronic social defeat stress (CSDS) paradigm. CSDS is an ethologically relevant paradigm that robustly and reproducibly induce depressive-like phenotypes and anxiety symptoms in rodents. We therefor performed CSDS on several cohorts of 3- and 6-month-old male mice and observed depressive-like behaviors using established measurements such as: nest building, food consumption, weight, sucrose preference testing and social interaction. From our behavior data, we found that both age groups demonstrated an acute stress response, social avoidance and significantly higher blood corticosterone levels. However, we found that anhedonic-like behavior was only observed in the 6-month animals, suggesting an age-dependent response to stress. A unique characteristic of the CSDS paradigm is that defeated animals are bimodally distributed into either â□□susceptibleâ□□ or â□□resilientâ□□ categories based on the social interaction test. Dot blot analysis revealed that in the cortex of susceptible animals, global levels of both 5mC and 5hmC were elevated compared to resilient animals. Using 5hmC-capture coupled with high-throughput sequencing, we analyzed the cortical 5hmC landscape in our 3-month CSDS resilient and susceptible animals. Differentially hydroxymethylated regions associated with either stress resilience or susceptibility were obtained and distinctive biological networks were identified. Cumulatively, our data demonstrates that the CSDS paradigm induces both physiological and biological symptoms of depression and that DNA modifications could play key roles in responding to stress that could ultimately contribute to mental illness.

PrgmNr 2495 - Transcriptome profiling using long-read sequencing to dissect the interplay between genetic variant and transcript variations in a population-based study

View session detail

Author Block: **A. Real**[1], C. Borel[2], N. Lykoskoufis[3], J. Seebach[1], A. Ramisch[4], E. T. Dermitzakis[3]; [1]Univ. Geneva, 1205, Switzerland, [2]Univ. of Geneva, Geneva 4, Switzerland, [3]Univ. of Geneva, Geneva, Switzerland, [4]Geneva, Switzerland

Disclosure Block: **A. Real:** None.

Over the last decade, RNA-seq technologies have vastly increased our knowledge of gene expression and transcript isoform signatures. However, current short-read RNA-seq methods show limitations in identifying complex transcript isoforms as full-length transcripts and RNA modifications are not retained. In this study, we address these limitations exploiting the advantages of long-read native poly(A) RNA-seq using the Oxford Nanopore Technologies (ONT) gridION platform. We direct RNA sequenced 60 genotyped lymphoblastoid cell lines from European ancestry obtaining a median of 3.2 million reads per sample with a median read length of 0.9M bps. We aim to understand the effect of genetic variation on transcription, alternative splicing and mRNA modification within a population.We identified 12,045 protein coding genes transcribed in at least 90% of the samples. In comparison, for the short-read RNA-seq (Illumina), the number of detected transcripts are similar (85%) with a variable transcripts abundance correlation among the samples (spearman r^2 0.49-0.68).We performed a preliminary isoform analysis on 42 samples using the FLAIR pipeline (Tang et al. Nature Communication, 2020), and detected a total of 234,513 transcript isoforms from which one third are expressed in at least 50% of the samples. It includes 40,449 isoforms accounting for 13,362 genes, 23080 are not annotated (GENCODE v27).Moreover, we observe variation of isoform usage across the population which is encouraging for our future analysis to understand the underlying mechanisms of these variations.Our next steps will be to perform allele-specific expression analysis to understand the effect of common and rare genetic variants on specific transcript alterations with the advantage of high resolution given by long-read sequence technology. Furthermore, the identification of RNA modifications combine with ChIP-seq data from our previous work will be of interest to dissect the contribution of RNA and histone modifications to gene regulation.

PrgmNr 2496 - Transcriptomic mapping of bursting kinetics in human cells

View session detail

Author Block: **B. Jin**, W. S. Bush; Case Western Reserve Univ., Cleveland, OH
Disclosure Block: **B. Jin:** None.

Most studies consider gene expression in terms of global transcript levels; however, growing evidence illustrates that promoters in eukaryotic cells tend to exhibit bursting patterns by repeatedly alternating between an activated and inactivated state. This promoter bursting can introduce significant cell-cell heterogeneity even in clonal cells. However, the molecular mechanism for bursting kinetics is not clear. Meanwhile, studying bursting kinetics has been limited to a small number of genes and model organisms due to difficulties in measuring transcript levels over time. Single-cell RNA-seq can measure individual cell expression profile, where each cell can be assumed as a temporal snapshot of an individual promoterâ□□s activity. Hence, we use single-cell RNA-seq data to at first time construct a transcriptomic kinetics map of two human cell lines: GM12878 and K562. Then, we further investigate how regulatory factors and histone markers are related to bursting kinetics within them. We found that the estimated bursting kinetics show two features: 1. Genes classified as bursting genes (14% genes for GM12878) exhibit a higher mean expression level. Furthermore, within a 2000bp window around the TSS, 4 out of 10 ranked regulators that discriminate bursting genes from non-bursting genes are shared between GM12878 and K562 (H3K36me3, POLR2Aphosphos2, CBFB, ZBED1) 2. The bursting frequency of genes in both GM12878 and K562 increase with the existence of putatively positive markers(e.g., H3K9ac, H3K27ac) and decrease trend with the existence of putatively negative markers(e.g., H3K9me3 and H3K36me3).By further examining experimental data perturbing enhancer regions within the K562 cell line, we validate the changes in promoter bursting kinetics. Overall, ~300 genes show a significant difference in the distribution of transcripts across perturbated cells compared to controlled cells. While some of these genes exhibit a similar increase or decrease in mean expression level, the changes in promoter kinetics are modulated differently. Our results show that bursting kinetics is independent information from the mean expression level and provides additional information on gene-specific regulatory mechanisms. Furthermore, as cell-cell heterogeneity has been associated with various cellular phenotypes(e.g., cell differentiation, aging, etc.), analyzing the bursting kinetics could add to our understanding of how regulatory variation influences disease risk.

PrgmNr 2497 - Chromolooper:Building genetic neighborhoods using a physically informed graph algorithm

View session detail

Author Block: **D. Borges-Rivera**, T. T. Tibbitts; Dogodan Therapeutics, Inc., Westford, MA
Disclosure Block: **D. Borges-Rivera:** None.
How a eukaryotic cell controls the many-to-many interactions between cis-regulatory elements and their cognate genes must be robust and simple, since it is a process that shows remarkable stability and fortitude to mutations while being rebuilt every cell division. Here we use a physically informed graph partitioning algorithm to calculate eukaryotic genetic neighborhoods, defining the genomic area by which each gene is likely to be regulated. We use as input chromatin loops, CTCF binding peaks, and RNA expression sourced from various large scale projects, such as ENCODE, to parametrize a modified louvain graph partitioning algorithm. We are able to recover previously described human genetic neighborhoods, such as the one encompassing IRX3 and IRX5, while predicting tens of thousands more per genome. These eukaryotic genetic neighborhoods, which reflect the local state-sensitive genetic regulatory architecture, are key for assessing the impact of disease-associated sequence variation not falling within protein coding regions. We connect our mendelian knowledge of genes to the non-coding using chromatin loops. We provide this algorithm as a python package, Chromolooper, to promote use in the wider community.

PrgmNr 2498 - Differential network analysis of time-series RNA-sequencing in response to perturbation

View session detail

Author Block: **S. Xue**[1,2], C. Piermarocchi[1], G. Mias[3]; [1]Dept. of Physics and Astronomy, Michigan State Univ., East Lansing, MI, [2]IQ Ctr., Michigan State Univ., East Lansing, MI, [3]Michigan State Univ., East Lansing, MI

Disclosure Block: **S. Xue:** None.

Differential Network (DN) analysis is a method that has long been used to interpret static gene expression data. The method identifies the rewiring of co-expression gene networks in response to external perturbations. Our study extends the DN method to time-series RNA sequencing (RNA-seq) data. We used DNs to analyze two sets of RNA-seq time series. The first dataset was obtained from an experiment profiling saliva expression in a human subject before and after vaccination with a 23-valent pneumococcal polysaccharide vaccine (PPSV23). Expression data covered 24 timepoints of hourly sampling over each of two periods, before and after vaccination respectively. The second dataset was from an ex-vivo experiment of primary B cell expression before and after treatment with Rituximab, covering six timepoints collected over 15 hours. We first preprocessed the experimental data to eliminate duplicate and sparsely sampled genes. We then calculated the fold-change between the treated and untreated data sets, selecting those genes which showed the biggest change in expression. With these target genes, we constructed a co-expression network for both the treated and untreated data sets. Subtracting the untreated network from the treated network produced a differential network. Finally, we detected gene communities and carried out Reactome pathway enrichment analyses to ascertain the biological significance of the DN results. Plotting the data from the DN into heatmaps enabled us to visualize the dynamic changes in individual communities. The DN method enabled us to identify biological pathways consistent with the mechanisms of action of the PPSV23 vaccine, and target pathways of Rituximab. The community detection algorithm on the DN revealed clusters of genes characterized by a collective behavior. Some DN communities, especially in the case of saliva, showed characteristic time signatures, outlining a chronological order in pathway activation in response to the perturbation. Our DN method confirmed the existence of temporal features in the changes of gene expression patterns. Moreover, we identified early and delayed responses within network modules in the saliva dataset, and three temporal trend patterns in the B cell data. Our DN method is simpler to implement and computationally more efficient than other network-based analysis for time-series RNA-seq data, such as Bayesian networks. Therefore, the DN method may be useful in drug development by providing a description of changes in gene expression in response to a drug.

PrgmNr 2499 - Effective gene expression prediction from sequence by integrating long-range interactions

View session detail

Author Block: **Z. Avsec**[1], V. Agarwal[2], D. Visentin[1], J. R. Ledsam[3], A. Grabska-Barwinskaâ[1], K. R. Taylor[1], Y. Assael[1], J. Jumper[1], P. Kohli[1], D. Kelley[4]; [1]DeepMind, London, United Kingdom, [2]South San Francisco, CA, [3]Google, Tokyo, Japan, [4]Calico Life Sci., South San Francisco, CA
Disclosure Block: **Z. Avsec:** Salary/Employment; DeepMind.
The next phase of genome biology research requires understanding how DNA sequence encodes phenotypes, from the molecular to organismal levels. How noncoding DNA determines gene expression in different cell types is a major unsolved problem, and critical downstream applications in human genetics depend on improved solutions. Here, we report substantially improved gene expression prediction accuracy from DNA sequence through the use of a new deep learning architecture called Enformer that is able to integrate long-range interactions (up to 100 kb away) in the genome. This improvement yielded more accurate variant effect predictions on gene expression for both natural genetic variants and saturation mutagenesis measured by massively parallel reporter assays. Notably, Enformer outperformed the best team on the critical assessment of genome interpretation (CAGI5) challenge for noncoding variant interpretation with no additional training. Furthermore, Enformer learned to predict enhancer-promoter interactions directly from DNA sequence competitively with methods that take direct experimental data as input. We expect that these advances will enable more effective fine-mapping of growing human disease associations to cell-type-specific gene regulatory mechanisms and provide a framework to interpret cis-regulatory evolution. To foster these downstream applications, we have made the pre-trained Enformer model openly available, and provide pre-computed effect predictions for all common variants in the 1000 Genomes dataset.

PrgmNr 2500 - Individual Specific Networks Hot Zone Detection

View session detail

Author Block: **B. Yousefi**[1,2,3], F. Melograna[3], D. Duroux[4], D. Qiao[5], Z. Li[3], P. J. Castaldi[5,6], C. P. Hersh[5,7], E. K. Silverman[5,7], M. H. Cho[5,7], B. Schwikowski[1], K. Van Steen[3,4]; [1]Systems Biology Group, Dept. of Computational Biology, Inst. Pasteur, Paris, France, [2]Sorbonne Univ., Complexite du vivant, Paris, France, [3]BIO3 â□□ Lab. for Systems Med., Katholieke Univ.it Leuven, Leuven, Belgium, [4]BIO3 â□□ Lab. for Systems Genetics, GIGA-R Med. Genomics, Univ. of LiÃ¨ge, LiÃ¨ge, Belgium, [5]Channing Div. of Network Med., Dept. of Med., Brigham and Womenâ□□s Hosp. and Harvard Med. Sch., Boston, MA, [6]Div. of Gen. Internal Med. and Primary Care, Dept. of Med., Brigham and Womenâ□□s Hosp. and Harvard Med. Sch., Boston, MA, [7]Div. of Pulmonary and Critical Care Med., Dept. of Med., Brigham and Womenâ□□s Hosp. and Harvard Med. Sch., Boston, MA
Disclosure Block: **B. Yousefi:** None.

Network approaches can provide unique pathobiological insights when applied to transcriptomics data. Particularly, gene expression correlation networks contain information about the regulatory relationships between genes and their co-functionality. A shortcoming is that interactions between genes may be individual-specific and can be underestimated in global network analysis approaches. To address this issue, Individual-specific networks (ISNs) have been proposed, thus allowing network analysis to capture inter-individual heterogeneity. ISNs can be constructed in multiple ways, and, in the literature, the same term identifies networks with very different rationales and construction methods. While ISNs may be powerful, their full potential has not yet been exploited: incorporating ISNs to construct complex predictors or explanatory variables in modelling or identifying latent classes of â□□similarâ□□ ISNs. Here, we focus on the method proposed in Kuijier et al. (2019), where the impact of removing an observation is captured by edge-specific weights.

We propose a novel algorithm (named HotZone) to find the network regions (â□□zonesâ□□) in ISNs that vary significantly across individuals and thus form interesting subgraphs for subsequent analyses and interpretation. Our algorithm consists of two steps: [i] find the gene communities in the global network, and [ii] test whether the individuals form sub-populations based on the variation of each gene module. The first step is performed using hierarchical clustering with a custom-defined resolution. Next, for each gene module, a graph-based individual similarity metric is constructed, which enables a statistical test to define the optimal number of clusters (related to the gene-module). Finally, after performing and correcting for multiple testing, the gene modules linked to sub-populations of individuals are recognized as network-hot-zones. We evaluate the performance of this algorithm on a reduced set of blood transcriptomics data from 2604 individuals in the COPDGene study, where the individuals have various levels of chronic lung disease.

Our results confirm the existence of biologically meaningful hot zones that foster simplified analyses with ISNs, and help improve our understanding mechanisms underlying patient-heterogeneity. Our analysis pipeline also introduces and highlights, in the field of ISN analysis, the concept of biclustering, for both gene expression and individuals. The code to perform the analysis will be freely available on GitHub.

PrgmNr 2501 - Local gene co-expression: prevalence and molecular features across human tissues and single-cells

View session detail

Author Block: **D. Ribeiro**[1,2], C. Ziyani[1], S. Rubinacci[1,2], R. Hofmeister[1,2], O. Delaneau[1,2]; [1]Univ. of Lausanne, Lausanne, Switzerland, [2]Swiss Inst. of Bioinformatics, Lausanne, Switzerland
Disclosure Block: **D. Ribeiro:** None.
Nearby genes are often expressed as a group. This local gene co-expression is more pronounced in the immediate vicinity of a gene (e.g. By leveraging gene expression measurements from the GTEx project across 49 human tissues and hundreds of individuals, we found local gene co-expression to be highly prevalent, occurring in 13% to 53% of genes per tissue. Indeed, we found that >64.000 gene pairs within 1Mb of each other are co-expressed in at least one tissue and that several features are associated with local gene co-expression, such as lower insulating CTCF sites between co-expressed gene pairs.
To understand how the observed local gene co-expression and its regulation manifests at the single-cell level, we analysed a public dataset of single-cell RNA-seq across 87 genotyped individuals in a specific cell type (iPSC, Cuomo et al. 2020 Nat. Comm.). By taking advantage of co-expression measurements across cells per individual, we identified hundreds of locally co-expressed gene pairs per individual. We found that these often participate in the same biological pathway and protein complex, indicating a strong functional link. Importantly, we show that local co-expression between two genes involves (i) concomitant co-transcription (GRO-seq data), (ii) co-regulation by shared enhancers (scATAC-seq data) and (iii) ultimately results in correlated protein levels (mass spectrometry data).
Finally, to understand whether genetic variation controls local gene co-expression, we identified expression quantitative trait loci (eQTLs) associating with co-expressed gene pairs. Interestingly, we found that these eQTLs are enriched in enhancer regions and are more often associated with multiple human traits than eQTLs associated with only one gene.
Our dissection of the genetic architecture of local gene co-expression through multiple technologies (RNA-seq, GRO-seq, ATAC-seq, proteomics) allows us to propose a model in which (i) the co-expression of nearby genes is prevalent and largely due to the sharing of enhancers, (ii) genetic variation perturbing the activity of shared enhancers may affect multiple genes and (iii) such variants have the potential to disrupt multiple traits and thus may contribute to the co-occurrence of diseases in the same individual.

PrgmNr 2502 - Modeling gene expression on the X chromosome for transcriptome-wide association studies

View session detail

Author Block: **X. Zhang**[1], W. S. Bush[2]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Case Western Reserve Univ, Cleveland, OH
Disclosure Block: **X. Zhang:** None.
The X chromosome is usually neglected in the analyses of genomic data. However, as we gain more knowledge of the disease associations and tissue-specific regulatory effects of autosomal loci, there is a growing interest in building prediction models of gene expression to understand the functions of non-coding variants on the X chromosome. A key challenge of X chromosome analysis is how to model copy number differences between males and females; genetic variants on the X chromosome are hemizygous for males, while females are heterozygous or homozygous diploid. Modeling is complicated by X inactivation when one of the two X chromosome copies of a female is permanently inactivated in nearly all somatic cells. Another challenge is two pseudoautosomal regions (PARs) on the X chromosome, on the tip of the short arm (Xp1) and the long arm (Xq28), respectively. Genes located in the PAR are inherited like autosomal genes in both genders, as there is a homologous region on the Y chromosome in males.
We explore modeling approaches using WGS and RNA-seq data in 13 different Brain tissues retrieved from the GTEx project, currently restricting analyses to males only. We used elastic-net regression, which combines LASSO and ridge penalty terms with a balanced LASSO-ridge penalty (mixing parameter alpha=0.5), as recommended by Wheeler et al. 2016. For each gene in each tissue, we tuned the penalization parameter *lambda* and assessed the model performance using a nested cross-validation strategy. There are 247 genes on chromosome band Xq28, with only approximately 100 genes expressed in each GTEx brain tissue. SNPs (MAF>0.05) within the 50k bp flanking windows of the gene were included in the model. We used mean squared error to determine the penalty parameter and select the best models. For example, we included 83 males to fit the prediction model of GAB3 Gene (ENSG00000160219, X: 153903529-153979858) in Frontal Cortex BA9. 2 out of 56 *cis*-SNPs were kept in the final model (*lambda*=0.36). The model R^2 is 0.0828, explaining 8.28% of the expression variance in males. These models estimate the genetically regulated gene expression on the X chromosome in males and can be readily used in Transcriptome-wide Association Studies.

PrgmNr 2503 - A signature of Neanderthal introgression on molecular mechanisms of environmental responses

View session detail

Author Block: **A. S. Findley**[1], X. Zhang[2], C. Boye[1], Y-L. Lin[3], C. Kalita[1], L. Barreiro[3], K. E. Lohmueller[2], R. Pique-Regi[4], F. Luca[4]; [1]Wayne State Univ., Detroit, MI, [2]Univ. of California, Los Angeles, Los Angeles, CA, [3]Univ. of Chicago, Chicago, IL, [4]Wayne State Univ, Detroit, MI
Disclosure Block: **A.S. Findley:** None.
Recent studies have suggested that introgression of archaic alleles in the genome of modern humans may have contributed to adaptation to environmental pressures such as pathogen exposure. Functional genomic studies have demonstrated that variation in gene expression across individuals and in response to environmental perturbations is a main mechanism underlying complex trait variation. We considered gene expression response to in vitro treatments as a molecular phenotype to identify genes and regulatory variants that may have played an important role in adaptations to local environments. We investigated if Neanderthal introgression in the human genome may contribute to the transcriptional response to environmental perturbations. We identified 27,349 introgressed Neanderthal alleles (N-SNPs) within the binding sites for 1,255 transcription factor (TF) motifs and computationally predicted that 73% alter TF binding. TFs enriched for N-SNPs which disrupt binding (e.g. *SWI5*, *MAFA*, and *ARNT*) are important for environmental responses, including ionizing radiation and hypoxia, and for glucose metabolism. Indeed, N-SNPs are enriched among SNPs that significantly affect TF binding in a SNP-SELEX dataset of type-2 diabetes associated loci. In particular, N-SNPs alter binding for 12 glucose metabolism-associated TFs (KS test p=0.036), including *IRF5*, *HOXA6*, *POU2F3*, and *PAX2*. We then considered eQTLs for genes differentially expressed in a panel of 52 cellular environments, resulting from 5 cell types and 26 treatments. We identify an enrichment for N-SNPs among eQTLs for genes differentially expressed in response to 8 treatments, including glucocorticoids, caffeine, and vitamin D. 132 genes which respond to a treatment were regulated by N-SNPs which show evidence of adaptive introgression. Finally, using massively parallel reporter assay data, we validated the regulatory function of 21 introgressed N-SNPs in the human genome, corresponding to 8 eQTLs regulating 15 genes that respond to environmental perturbations. These findings expand the set of environments where archaic introgression may have contributed to adaptations to local environments in modern humans and provide experimental validation for the regulatory function of introgressed variants.

PrgmNr 2504 - Allelic frequency comparison to improve epidemiological decisions on hereditary diseases in a Latino population

View session detail

Author Block: **A. Sanchez-Gomez**, L. J. Moreno-Giraldo, J. M. Satizabal-Soto; Univ. del Valle, Cali, Colombia
Disclosure Block: **A. Sanchez-Gomez:** None.
Normal 0 false false false EN-US X-NONE X-NONE /* Style Definitions */ table.MsoNormalTable {mso-style-name:'Tabla normal'; mso-tstyle-rowband-size:0; mso-tstyle-colband-size:0; mso-style-noshow:yes; mso-style-priority:99; mso-style-parent:''; mso-padding-alt:0in 5.4pt 0in 5.4pt; mso-para-margin-top:0in; mso-para-margin-right:0in; mso-para-margin-bottom:8.0pt; mso-para-margin-left:0in; line-height:107%; mso-pagination:widow-orphan; font-size:11.0pt; font-family:'Calibri',sans-serif; mso-ascii-font-family:Calibri; mso-ascii-theme-font:minor-latin; mso-hansi-font-family:Calibri; mso-hansi-theme-font:minor-latin; mso-bidi-font-family:'Times New Roman'; mso-bidi-theme-font:minor-bidi;}
The allelic frequency of sequence variants with clinical value for the diagnosis of hereditary diseases is a valuable indicator when making decisions for the epidemiological surveillance of these diseases in populations at risk. However, most of the values â□□â□□recorded in the available databases, for populations other than Caucasians and Afro-descendants, may present sampling biases in the values â□□â□□of this indicator. For this reason, our working group on Hunter's disease (MPS II) carried out a sampling in a disease-free population originating from the southwestern part of Colombia, in order to calculate the allelic frequencies of the main circulating sequence variants associated with the gene IDS and compare these values â□□â□□with those registered in the Gnomad database for those same variants in the Latino/admixture American subgroup. Our findings show a statistically significant difference between the values â□□â□□recorded in Gnomad and those observed in the study population. This difference shows an underestimation of the allele frequencies studied at the Gnomad base. When extrapolating the allelic frequency values â□□â□□with the incidence of MPS II in the geographical region of the study, a greater concordance of the epidemiological data with the population data is seen. Based on the values â□□â□□observed in the study population, it is suggested to maintain an active search process for undiagnosed patients to avoid clinical under-registration and thus improve the possibilities of timely care with therapeutic treatments available on the market.

PrgmNr 2507 - Characterizing the evolutionary history of HLA genes in sub-Saharan African populations

View session detail

Author Block: **D. N. Harris**[1], M. McQuillan[2], O. Ozur[3,4], M. Hansen[2], S. Fan[5], A. Ranciaro[2], W. R. Beggs[2], T. Nyambo[6], S. Mpoloka[7], G. Mokone[8], D. Woldemeskel[9], G. Belay[9], C. Fokunang[10], A. K. Njamnshi[11], J. Duke[12], D. Ferriola[12], M. Ntinou[12], I. Pagkrati[12], G. Damianos[12], T. Mosbruger[12], T. L. Lenz[4,13], D. Monos[14,12], S. A. Tishkoff[15,16]; [1]Genetics, Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Res. Unit for Evolutionary Immunogenomics, Dept. of Biology, Univ. of Hamburg, Hamburg, Germany, [4]Res. Group for Evolutionary Immunogenomics, Max Planck Inst. for Evolutionary Biology, Ploen, Germany, [5]State Key Lab. of Genetic Engineering, Human Phenome Inst., Sch. of Life Sci., Fudan Univ.,, Shanghai, China, [6]Kampala Intl. Univ. in Tanzania (KIUT) P.O.Box 9790, Dar es Salaam, Tanzania, United Republic of, [7]Univ. of Botswana, Biological Sci., Gaborone, Botswana, [8]Univ. of Botswana, Faculty of Med., Gaborone, Botswana, [9]Dept. of Biology, Addis Ababa Univ., Addis Ababa, Ethiopia, [10]Dept. of Pharmacotoxicology and Pharmacokinetics, Faculty of Med. and BioMed. Sci., The Univ. of YaoundÃ© I, YaoundÃ©, Cameroon, [11]Dept. of Neurology, Central Hosp. YaoundÃ©; Brain Res. Africa Initiative (BRAIN), NeuroSci. Lab, Faculty of Med. and BioMed. Sci., The Univ. of YaoundÃ© I, YaoundÃ©, Cameroon, [12]Dept. of Pathology and Lab. Med., Childrenâ□□s Hosp. of Philadelphia, Philadelphia, PA, [13]Res. Unit for Evolutionary Immunogenomics, Dept. of Biology, Univ. of Hamburg, Hamburg, Germany, [14]Dept. of Pathology and Lab. Med., Perelman Sch.s of Med., Univ. of Pennsylvania, Philadelphia, PA, [15]Univ. of Pennsylvania Sch. of Med., Philadelphia, PA, [16]Dept. of Biology, Univ. of Pennsylvania, Philadelphia, PA

Disclosure Block: **D.N. Harris:** None.

The Human Leukocyte Antigen (HLA) gene complex encodes proteins that control antigen-specific immune responses. Variants in the HLA gene region are associated with many diseases associated with inflammation, such as Type 1 Diabetes and Multiple Sclerosis. Africans are underrepresented in genetic studies; therefore, it is imperative to characterize genetic variation of the HLA gene region in African ancestry genomes. We sampled ethnically and genetically diverse populations with varying subsistence patterns and infectious disease exposures from four African countries. Targeted sequencing of the HLA genes in 489 individuals generated 4,506 consensus sequences, and among those sequences there were 754 unique alleles. Among these alleles there were 348 sequences that were either novel or that completed the genomic sequences of incompletely published alleles. For a subset of 180 individuals, we generated high coverage whole genome sequence data that enables us to compare genetic variation in the HLA region with the rest of the genome. Heterozygosity is greatly increased within the HLA region compared to genome wide levels, and we estimate that variants in the classical HLA genes are some of the oldest in the human genome. Constructing a phylogeny from genome-wide patterns of genetic variation demonstrates genetic differentiation that correlates with language and geography. By contrast, we do not observe the same pattern in the phylogeny constructed from genetic variation in HLA genes, where individuals do nut cluster by geographic location or language group. This observation, combined with the increased heterozygosity and prevalence of old variants, is consistent with balancing selection at this region. To identify signatures of balancing selection and local adaptation we scanned the HLA region using the Di statistic, which identifies variants that are strongly differentiated in focal populations, and the Tajimaâ□□s D statistic which is based on the allele frequency distribution. Tajimaâ□□s D statistic identifies a strong signal of balancing selection across the HLA region. We also identify multiple candidates of local adaptation in hunter-gatherer populations, possibly due to differential pathogen exposure. The top selection signals are in the *HLA-B* and *TAP1* genes, the latter of which facilitates the transporting and loading of peptides to the HLA. Future functional genomics data will demonstrate how these positively selected

variants impact immune response and their role in human diseases. Supported by NIH grants R35GM134957-01, 5T32DK007314-39, ADA grant 1-19-VSN-02 and funding from the Chan Zuckerberg Institute.

PrgmNr 2508 - Deep variational autoencoders for population genetics: applications in classification, imputation, dimensionality reduction, and novel lossless data compression

View session detail

Author Block: **M. Geleta**[1,2], D. Mas Montserrat[3], C. D. Bustamante[3], X. GirÃ³-i-Nieto[2], A. Ioannidis[3,4]; [1]Univ. of California Irvine, Irvine, CA, [2]Univ.t PolitÃ¨cnica de Catalunya, Barcelona, Spain, [3]Dept. of BioMed. Data Sci., Stanford Univ. Sch. of Med., Palo Alto, CA, [4]Inst. for Computational & Mathematical Engineering, Stanford Univ. Sch. of Engineering, Palo Alto, CA
Disclosure Block: **M. Geleta:** None.
In this study we show the power of variational autoencoders (VAEs) for a variety of tasks relating to the interpretation and compression of genomic data. The unsupervised setting allows for detecting and learning of granular population structure and inferring of new informative latent factors, opening up an avenue for applications in dimensionality reduction, data simulation, population classification, imputation, and lossless genomic data compression. The latent spaces of VAEs are able to capture and represent clearly differentiated Gaussian-like clusters of similar genetic composition on a fine-scale with a relatively small number of SNPs as input. Furthermore, sequences can be decomposed into latent representations and reconstruction errors (residuals) providing a sparse representation that provides a means for efficient lossless compression.

Identifying genetic clusters can be important when performing genome-wide association studies and provides an alternative to self-reported ethnic labels, which are culturally constructed and vary according to the location and individual. A variety of unsupervised dimensionality reduction methods have been explored in the past for such applications, including PCA, MDS, t-SNE, and UMAP. Our proposed VAE can represent the population structure as a Gaussian-distributed continuous multi-dimensional representation and as classification probabilities providing flexible and interpretable population descriptors.

We train our VAE method with several worldwide whole genome datasets from both humans and canids and evaluate the performance of the different proposed applications with networks with and without ancestry conditioning. Our experiments show that different population groups have significantly differentiated compression ratios and classification accuracies. Additionally, we analyze the entropy of the SNP data, noting its effect on compression across populations and connect these patterns to historical migrations and ancestral relationships.

PrgmNr 2509 - Estimating ancestry-specific allele frequencies in admixed samples from large-scale sequencing data

View session detail

Author Block: **J. Wang**, A. V. Smith, The Trans-Omics for Precision Medicine (TOPMed) Consortium, G. Abecasis, S. ZÃ¶llner; Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **J. Wang:** None.
Estimates of allele frequency distributions across different populations provide important insights into the transferability of GWAS results, population specific drivers of disease etiologies, and population genetic estimates such as natural selection. Allele frequencies from recently admixed groups are often more difficult to interpret in this context and it is of great interest to estimate the allele frequencies of the underlying ancestral groups. We leverage local ancestry to estimate ancestry-specific allele frequencies. After local ancestry is inferred, the ancestry of variants with homozygous ancestry background can be directly observed. To accommodate variants with heterozygous ancestry background, we developed two algorithms: (1) We derive an Expectation-Maximization approach that treats the phase information in individuals with heterozygous ancestry background as augmented data to incorporate additional information like sample-wide ancestry distributions to inform ancestral composition at the variants with heterozygous background. (2) We leverage statistical phasing of samples by duplicating each haplotype and pairing the two copies into one pseudo diploid that is homozygous for local ancestry everywhere. Both algorithms are capable of interpreting large-scale sequencing data and multi-way admixture patterns. Simulation results show a small bias of the EM estimates of 0.01 and effective sample size > 100. Variance of the EM estimator decreases with larger sample size and corresponding ancestry proportion. We plan to apply both algorithms to genome sequencing data from the TOPMed Consortium, and estimate allele frequencies based on samples. We have inferred local ancestry in this sample observing of ~55% European haplotypes, ~25% African haplotypes, ~5% Native American haplotypes, ~5% Central and South Asian haplotypes, ~5% East Asian haplotypes, ~4% Middle Eastern haplotypes, and 0.99 for common variants with allele frequency greater than 0.1%. We estimate differentiated allele frequency patterns among different ancestral groups, and observe how similarity in allele frequencies across all frequency bins > 0.1% are closely related with population histories.

PrgmNr 2510 - Exploring the role of local sequence context on patterns of germline mutation

View session detail

Author Block: **A. Beck**[1], S. Zollner[2], J. Li[2]; [1]Univ. of Michigan Sch. of Publ. Hlth., Ann Arbor, MI, [2]Univ Michigan, Ann Arbor, MI
Disclosure Block: **A. Beck:** None.
Mutation rates are known to vary across the genome, and while this variation has been shown to be correlated with both local sequence context and other genomic and epigenomic features, our current understanding of how these features influence mutation rates and processes in the germline is limited. Matrix factorization approaches have been utilized in cancer genomics to generate signatures of potential mutational processes, which have then been linked to well-characterized biological processes. However, these methods are limited to considering only one or two flanking bases up- and down-stream of a mutation due to the exponential growth of the number of motifs as more flanking positions are considered. In order to develop methods to study processes underlying germline mutation, an understanding of the extent to which nucleotides at positions flanking a variant site influence the rate of mutation in the germline, and to what degree interactions between bases at flanking positions influence the rate is an important first step. Here we evaluate the relationship between local sequence context and the distribution of singletons across 2,504 unrelated individuals in the 1000 Genomes high-coverage data. We find that models which control for genomic location (and thus genomic features) perform better than models which only account for genome-wide nucleotide composition. When evaluating the marginal influence of nucleotides at positions flanking a variant site, nucleotides at positions beyond the +/- 1 bp positions considered by signature decomposition approaches are shown to be correlated with mutation rate variation. We also find that interactions between nucleotides are also associated with mutation rate variation, but that their influence decreases as we consider both interactions between positions further from each other and higher-order interactions. These patterns are observed across all populations within 1000 Genomes. Taken together, our results suggest that methods for uncovering underlying signatures of mutation processes in the germline could benefit from both controlling for the influence of genomic positions and incorporating more sequence context without accounting for higher-order interactions, preventing the exponential growth of the complexity of the model as more flanking positions are considered.

PrgmNr 2511 - Genetic and pharmacological causes of germline hypermutation

View session detail

Author Block: J. Kaplanis[1], B. Ide[2], R. Sanghvi[3], M. Neville[4], P. Danecek[5], E. Prigmore[1], P. J. Short[6], G. Gallone[1], J. McRae[7], C. Odhams[8], L. Moutsianas[9], Genomics England Research Consortium, J. Carmichael[10], A. Barnicoat[11], H. Firth[1], P. O'Brien[2], R. Rahbari[12], M. E. Hurles[1]; [1]Wellcome Sanger Inst., Cambridge, United Kingdom, [2]Univ. of Michigan, Ann Arbor, MI, [3]Wellcome Sanger Inst., CAMBRIDGE, United Kingdom, [4]Sanger Inst., Cambridge, United Kingdom, [5]Wellcome Sanger Inst., Hinxton, United Kingdom, [6]Wellcome Trust Sanger Inst, Hinxton, United Kingdom, [7]Wellcome Trust Sanger Inst, Hinxton, Cambridgeshire, United Kingdom, [8]Loughton, United Kingdom, [9]Genomics England, London, United Kingdom, [10]Cambridge Univ. Hosp., Cambridge, United Kingdom, [11]Great Ormond Street Hosp., London, United Kingdom, [12]Wellcome Trust Sanger Inst., Cambridge, Cambridgeshire, United Kingdom
Disclosure Block: J. Kaplanis: None.

Mutation in the germline is the source of all evolutionary genetic variation and a cause of genetic disease. Previous studies have shown parental age to be the primary determinant of the number of new germline mutations seen in an individualâ□□s genome.

Here we analysed the genome-wide sequences of 21,879 families with rare genetic diseases from the 100,000 Genomes Project (100kGP) and the Deciphering Developmental Disorders (DDD) Study and identified 12 hypermutated individuals with 2-7 times more de novo single nucleotide variants (dnSNVs) than expected. For 2/12 hypermutated individuals, the excess mutations appeared to have occurred post-zygotically, however for the majority (9) of these hypermutated individuals, the excess dnSNVs phased paternally implicating the father as the source of hypermutation. For 5 of these fathers, characteristic mutational signatures and clinical records of cancer treatment prior to conception strongly implicated the mutagenic influence of two different classes of chemotherapeutics: platinum-based drugs (3 families) and mustard-derived alkylating agents (2 families). We also identified likely paternal mutator variants in two hypermutated families. These were rare homozygous missense variants in known DNA repair genes: *XPC* and *MPG*. Functional and clinical data supported the mutagenic nature of these variants.

We estimated that germline hypermutation accounted for 7% of the variance in germline mutation rate in the 100kGP cohort and parental age explained ~70% which is substantially smaller than previous estimates. For the residual ~20%, we found that rare variants in known DNA repair genes are unlikely to account for a large proportion of this unexplained variance and that polygenic contributions from common variants (MAF>1%) are unlikely to make a large contribution although the polygenic contribution of intermediate frequency paternal variants (0.0011.5) fold increase in mutation rate.

Our results suggest that the germline is well protected from mutagenic effects, hypermutation is rare and relatively modest in degree. The absolute risk of germline hypermutated individuals having a rare genetic disease is low; we anticipate most will not be affected and germline hypermutation will also be observed in healthy population cohorts.

This research was made possible through access to the data and findings generated by the 100,000 Genomes Project http://www.genomicsengland.co.uk

PrgmNr 2512 - Genetic structure, kinship, and ancient pathogen diversity in a Moche tomb in Northern Peru

View session detail

Author Block: **C. Guerra Amorim**[1,2], S. Neuenschwander[2], Y. Arizmendi-CÃirdenas[2], B. Mota[2], D. I. Cruz-DÃivalos[2], C. Barbieri[3], J. Verano[4], L. Fehren-Schmitz[5], A. Malaspinas[2]; [1]California State Univ. Northridge, Los Angeles, CA, [2]Univ. of Lausanne, Lausanne, Switzerland, [3]Univ. of Zurich, zurich, Switzerland, [4]Tulane Univ., New Orleans, LA, [5]Univ. of California Santa Cruz, Santa Cruz, CA
Disclosure Block: **C. Guerra Amorim:** None.
The Moche culture flourished in the Northern Peruvian coast between the years 100 BCE and 800 CE, extending over a region of 500 km at its apogee. The Moche engineered many elaborate metal artifacts and figurative pottery, possessed a remarkably complex social structure, and presented dynamic funerary practice. *Huaca Cao Viejo* is one of the most prominent Moche archeological sites. It is located within the *El Brujo* archeological complex, in Northern Peru, and it is mostly known for the undisturbed burial of a female leader known as the Lady of Cao, discovered in 2006. Here, we analyze novel whole-genome sequences of n=7 ancient individuals, dated to ~350-600 CE, buried in a tomb adjacent to the main Lady of Cao burial. Our analysis of the genome-wide diversity of these ancient individuals shows that they are most similar to present-day Native American individuals from the same geographical region where they were buried (i.e., the Northern Peruvian coast). This suggests that no major population turnovers happened in the region from ~500 CE until the period of contact (~1,500 CE). The practice of burying additional persons (referred to in the literature as â□□retainersâ□□) alongside a principal individual is a common feature in the Moche mortuary practice. Yet, the relationship between retainers and principal individuals remains elusive, with no conclusive indications whether they were unrelated servants or biologically related family members. To shed light on this issue, we generated additional ancient DNA data for n=3 ancient individuals: the Lady of Cao and two individuals buried in her tomb. We found that the Lady of Cao was biologically related (2nd to 4th-degree relatedness) to the latter two individuals, but not to any of the â□□retainersâ□□ in the adjacent tomb. Furthermore, we screened all the sequenced genomic libraries for known human viruses and detected the presence of ancient Hepatitis B virus (HBV) DNA molecules in one individual.

PrgmNr 2513 - Population genetic analyses implicate biogenesis of translation machinery in human ageing

View session detail

Author Block: **S. Javidnia**[1], S. Cranwell[1], S. H. Mueller[1], K. B. Kuchenbaecker[2], N. Alic[1]; [1]UCL, London, United Kingdom, [2]Univ Coll. London, London, United Kingdom
Disclosure Block: **S. Javidnia:** None.
Reduced provision of protein translation machinery promotes healthy ageing in a number of animal models. In humans, however, inborn impairments in translation machinery are a known cause of a range of developmental disorders, collectively called ribosomopathies. Here, we employ population genetic approaches to investigate if adult, tissue-specific biogenesis of translation machinery drives human ageing. We assess naturally occurring variation in the expression of genes encoding subunits specific to the two RNA polymerases (Pols) that transcribe ribosomal and transfer RNAs, namely Pol I and III, and the variation in expression of ribosomal protein (RP) genes, using Mendelian Randomisation. We find each causally associated with human longevity (Ã=-0.15Â±0.047, p=9.6x10-4; Ã=-0.13Â±0.040, p=1.4x10- 3; Ã=-0.048Â±0.016, p=3.5x10-3, respectively). These associations do not appear to be mediated by altered susceptibility to a single disease. Interestingly, we find that reduced expression of Pol III, RPs or Pol I promote longevity from different organs, namely visceral adipose, liver and skeletal muscle, echoing the tissue-specificity of ribosomopathies, and we provide evidence that Pol I and RPs may act from organs where their expression is limiting. Our study demonstrates the utility of leveraging genetic variation in expression to elucidate how essential cellular processes impact human ageing. Our findings extend the evolutionary conservation of protein synthesis as a process that drives animal ageing to include humans.

PrgmNr 2514 - Revealing the recent demographic history of Europe via haplotype sharing in the UK Biobank

View session detail

Author Block: **E. H. Gilbert**[1,2], G. L. Cavalleri[1,2]; [1]Royal Coll. of Surgeons in Ireland, Dublin, Ireland, [2]FutureNeuro SFI Res. Ctr., RCSI, Dublin, Ireland
Disclosure Block: **E.H. Gilbert:** None.
Haplotype-based analyses have recently been leveraged to interrogate fine-scale structure in specific geographic regions, notably in Europe. An equivalent understanding across the whole of Europe with these tools however is lacking and would provide an updated map of the European genetic landscape. Similarly, a study of Identity-by-Descent (IBD) sharing in a large sample of pan-Europe genotypes would allow both direct comparison between different demographic histories and in parallel identify communities conducive to genetic mapping. In this context, we sought to investigate the extent of European ancestry captured in the UK Biobank (UKBB), a large genetic dataset with world-wide ancestry. We sampled 4,920 UKBB individuals with a European birthplace and investigated population structure and demographic history in Europe. With one of the largest samples of genotypes from across the geographical extent of Europe we show in parallel the variety of footprints of demographic history in different genetic regions around Europe and expand knowledge of the genetic landscape of the east and south-east of Europe. We highlight novel analysis of island populations such as Malta and the Channel Islands, demonstrating with IBD-segment sharing the extent of population isolation and size. Our work builds and expands upon previous work in Europe and specific populations, highlighting UK Biobank as a source of diverse ancestries beyond Britain. We find novel results in multiple communities in Europe that are of interest to genetic mapping.

PrgmNr 2515 - Sequencing the general population at a fine-geographic scale: the French POPGEN project

View session detail

Author Block: **A. Herzig**[1], G. Le Folgoc[1], M. Goldberg[2], M. Zins[2], E. GÃ©nin[3]; [1]Inserm, Univ Brest, EFS, UMR 1078, GGB, Brest, France, [2]Inserm-Paris Saclay Univ. UMS 011, UniversitÃ© de Paris, Villejuif, France, [3]Inserm UMR1078, Brest, France
Disclosure Block: **A. Herzig:** None.
Population-level genome-sequencing projects have been undertaken by many nations. The availability of these, often immense, datasets has accelerated methodological development and provided new avenues for genetic epidemiological investigation. France will soon join the growing list of nations where such datasets have been compiled through the POPGEN project; a pilot study of the National Genomic plan led by Inserm. Whilst the envisaged dataset may not rank highest in terms of size, it promises to break new ground with its meticulous selection criteria.
Using questionnaires completed by volunteers from the Constances cohort, we selected participants based on the birth-places of their four grand-parents. As participants are adults over 35 years of age (mean of 59, range of 36-80), the average birth year of their grandparents was 1906. Our key assumption was that an individual with all four grand-parents born within a small locality should be a reliable representative of that region. By using such a selection criterion, we will obtain insights on the genetic diversity between French regions at the start of the 20th century when population movement remained limited. We selected 15,000 individuals respecting a geographic distribution based on historical population proportions in France from the 1901 census. Hence, POPGEN will capture a complete genetic cross-section of the French population. The connection between each participant and a geographical origin will permit a high resolution for analyses of genetic fine-structure.
The selected individuals were sent saliva collection kits through the post, a design aimed to best avoid recruitment bias and maximize participation. A pilot study successfully showed the appropriateness of this DNA collection method for whole-genome sequencing (WGS) as well as the possibility for parallel sequencing of salivary microbiomes. Using a smaller but similarly selected group of 856 French individuals with WGS data, we established the potential for using POPGEN as a population specific reference panel for imputation. Finally, we present analyses of demographic patterns observable through the multi-generational birth-place data and how we predict these could manifest in the genetic dataset that will materialize in 2022.

PrgmNr 2516 - SIRPA V1/V2 haplotypes were selected by balancing selection

View session detail

Author Block: **N. Vince**, F. Haspot, S. Le Bas - Bernardet, N. Poirier, P-A. Gourraud, G. Blancho, S. Limou; UniversitÃ© de Nantes, Nantes Cedex 01, France
Disclosure Block: **N. Vince:** None.
Signal regulatory proteins (SIRPs) are a group of cell surface receptors involved in immune functions through immunoglobulin-like domains. Human SIRPs are composed of 5 different proteins (Î±, Î²1, Î²2, É£, Î´), but only SIRP was extensively studied so far. The SIRPÎ±/CD47 ligand/receptor pair is under active therapeutic research to tame the â□□donâ□□t eat meâ□□ mechanism preventing CD47+ cell phagocytosis, with applications in transplantation (signal stimulation) or oncology and infectious diseases (signal blocking). Both therapeutic strategies may be sensitive to polymorphisms, as *SIRPA* is highly polymorphic with 2 main haplotypes described in humans, V1 and V2, translating to 2 different proteins (13 amino acids differences). Indeed, monoclonal antibodies may bind these haplotypes with different affinity. Our study aimed at characterizing the evolutionary history of the *SIRPA* polymorphisms.
We investigated the potential selection signature on the *SIRPA* V1/V2 polymorphisms using the 1000 Genomes project 30X dataset. We computed the fixation index (F_{ST}; degree of differentiation between two populations), extended haplotype homozygosity (EHH; haplotype length which may detect recent selection event) and Tajimaâ□□s D (neutrality test, positivity may reflect balancing selection).
We showed that the V1/V2 *SIRPA* haplotypes are present at high frequencies across all human populations with large disparities (V1 frequencies in 3 populations: 71% in African, 63% in European and 31% in East-Asian). However, these distribution differences did not translate into high F_{ST} (below 0.04). V1 and V2 haplotypes also showed similar length across populations, reflecting the absence of recent selection event. In all 3 tested populations, the highest Tajimaâ□□s D score across the whole chromosome 20 was found within *SIRPA* (5.1 in African, 5.2 in European and 4.6 in East-Asian).
The combination of high Tajimaâ□□s D and high frequencies in all 3 populations tested suggest that *SIRPA* V1/V2 haplotypes were, and might still be, under balancing selection. This means that, during the course of evolution, 2 SIRPÎ± versions expressed in a given individual (i.e. V1/V2 heterozygous state) could have been important to prevent an inhibition of the immune response against infection. Differential distribution of *SIRPA* haplotypes V1/V2 frequencies across Human population may impact therapeutic successes (such as monoclonal antibodies) and individual *SIRPA* genotyping might be useful before selecting a treatment targeting SIRPÎ±.

PrgmNr 2517 - ï»¿The genetic landscape of South Native American populations

View session detail

Author Block: **M. Castro e Silva**[1], C. M. Couto-Silva[1], R. B. Lemes[1], K. Nunes[1], D. Comas[2], T. HÃ¼nemeier[1]; [1]Univ. of SÃ£o Paulo, SÃ£o Paulo, Brazil, [2]Univ.t Pompeu Fabra, Barcelona, Spain
Disclosure Block: **M. Castro e Silva:** None.
South America presents one of the most diverse sets of environments in the world, which has influenced human biological and cultural evolution from the very beginning of its occupation. At the same time, these environments were modified and adapted to human survival. This dynamic process generated a vast cultural diversity, contrasting with the low levels of Native American genetic diversity described hitherto. We analyzed genomic data from 383 present-day natives from more than 50 ethnolinguistic groups genotyped with the Affymetrix Human Origins array to shed light on the South America, Mesoamerica and northern Mexico genetic landscape. First, we find no evidence of a sharp genetic divide between Andeans and Amazonians. Conversely, we demonstrate that the genetic variation, as well as the homozygosity level, is correlated with longitude, which is compatible with an isolation by distance model, possibly tracing back to an initial settlement from the Pacific coast. Noteworthy, present-day Native American diversity recapitulates ancient local ancestries. Demographic inferences indicate higher population sizes and resilience of western South American groups to the population collapses caused by the European invasion and also indicate some pre-contact demic expansions. Finally, we find strong evidence of cultural exchanges that led to a language replacement in western Amazonia in pre-contact times.

PrgmNr 2518 - The GLAD Initiative: a database for supporting population genetics and association studies in the Americas

View session detail

Author Block: **V. Borda**[1], D. Loesch[2], B. Guo[3], R. Laboulaye[1], J. French-Kwawu[4], D. Veliz[1], GLAD Invertigators, TOPMed Population Genetics Working Group, E. E. Perez-Palma[5], I. Mata[6], R. D. Hernandez[7], E. M. Tarazona-Santos[8], T. D. O'Connor[1]; [1]Univ. of Maryland Sch. of Med., Baltimore, MD, [2]Univ. of Maryland Baltimore, Baltimore, MD, [3]China, [4]Univ. of Maryland Baltimore, Pikesville, MD, [5]Univ. zu KÃ¶ln, KÃ¶ln, Germany, [6]Lerner Res. Inst., Genomic Med., Cleveland Clinic, Cleveland, OH, [7]San Francisco, CA, [8]Univ Federal Minas Gerais, Belo Horizonte, Brazil

Disclosure Block: **V. Borda:** None.

Background: Latin Americans and U.S. Hispanics (LA) are an interesting model for evolutionary and medical studies as they result from recent admixture of highly differentiated sources (Native Americans, Europeans, East Asians, and Africans). However, underrepresentation of genetic studies in these heterogeneous groups does not allow for a comprehensive understanding of genetic determinants underlying complex traits. To overcome the underrepresentation issue, we built a comprehensive collection of genome-wide information of LA in the Genetics of Latin American Diversity (GLAD) database. **Structure of GLAD:**By gleaning all possible LA individuals through dbGaP and whole genome sequencing projects across Latin America, we gathered 80,542 (26,289 WG sequenced) LA individuals that are either self-described or Admixture-defined (i.e. individuals with at least 2% Native American ancestry). After data curation, imputation with the TOPMed Imputation Server, and merging, the GLAD data included 2.1M post QC SNPs with an $r^2 > 0.9$ (6.2M with an $r^2 > 0.3$) spanning at least sixteen countries in the Americas. By joining local ancestry and Identity-By Descent (IBD) analyses on GLAD data, we tracked patterns of higher IBD sharing on the Pacific and Atlantic sides of the continent related to Native American and African ancestries, respectively. These patterns reflect the dynamics of both ancestries during colonial times. **GLAD as a Resource:**It is a big challenge to obtain genetic information for a large number of control individuals for a specific phenotype and ancestry, specially for the LA. This is aggravated by the fact that most large-scale genetic studies suffer from data sharing issues due to consent or difficulty of access through public databases. To overcome this problem, we will provide support for association studies of LA by sharing summary statistics generated from the GLAD database in order to increase the power of association and admixture mapping studies without transferring individual-level data. We will identify control subject of similar genetic background to cases submitted by external users using a matching algorithm based on summary statistics of ancestry and genetic background (e.g. PCA and admixture proportions). Finally, our GLAD database contributes to the understanding of the genetic architecture of traits as well as the fine-scale patterns of Latin American genetic diversity.

PrgmNr 2519 - *Genetics Adviser: The development, usability and acceptance testing of a patient-centered digital health application to support clinical genomic testing*

View session detail

Author Block: **M. Clausen**[1], R. Kodida[2], S. Shickh[3], E. Reble[4], C. Mighton[5], D. Hirjikaka[2], J. Sam[6], S. Krishnapillai[2], E. Adi-Wauran[7], G. Feldman[2], E. A. Glogowski[8], S. Shastri-Estrada[2], A. Scheer[2], E. Seto[9], C. T. Shuman[10], N. Baxter[11], A. Eisen[12], C. Elser[13], R. Kim[14], J. Lerner-Ellis[13], J. C. Carroll[15], K. A. Schrader[16], H. Faghfoury[17], Y. Bombard[15]; [1]209 Victoria Street, Toronto, ON, Canada, [2]St. Michael's Hosp., Toronto, ON, Canada, [3]St. Michael's Hosp. and Univ. of Toronto, Toronto, ON, Canada, [4]St. Michael's Hosp., Toronto, Canada, [5]St. Michael's Hosp. & Univ. of Toronto, Toronto, ON, Canada, [6]St. Michael's Hosp., North York, ON, Canada, [78]Mem Sloan Kettering Cancer Ctr, New York, NY, [9]Univ. of Toronto, Toronto, ON, Canada, [10]Hosp for Sick Children, Toronto, ON, Canada, [11]Univ. of Melbourne, Melbourne, Australia, [12]Sunnybrook Hosp., Toronto, ON, Canada, [13]Mount Sinai Hosp., Toronto, ON, Canada, [14]Univ. Hlth.Network/Mt Sinai Hosp., Toronto, ON, Canada, [15]Univ of Toronto, Toronto, ON, Canada, [16]Univ British Columbia, Vancouver, Canada, [17]Fred a Litwin and family Ctr. in Genetic Med., Toronto, ON, Canada

Disclosure Block: **M. Clausen:** None.

Background: Increasing demand for genomic testing coupled with existing workforce shortages in clinical genetics has placed unsustainable pressure on the standard models of care. Patient-facing digital health applications can empower patients and provide sustainable and scalable clinical solutions to address this gap. **Aim:** To transform our original Genomics ADvISER decision aid into a comprehensive patient-centered digital health application that will deliver education, counseling, and return of results for patients undergoing various form of genomic testing. **Methods:** Driven by user-centred design principles, we developed and conducted usability and acceptance testing of the application using an iterative, mixed-methods process consisting of: 1) consultations with an advisory board of providers and patients; 2) analysis of qualitative interviews with prior patients who used the original Genomics ADvISER decision aid; 3) creation of a digital wireframe prototype; 4) usability testing of the prototype; and 5) acceptability testing of the final digital application. ***Results:*** **Prototype development:** We created a new digital application, called the â□□Genetics Adviserâ□□, building on our original â□□Genomics ADvISERâ□□ that incorporated feedback from our advisory board, patients, genetics experts and the general public. The Genetics Adviser is designed to be easily adaptable to the needs of different types of patients, test modalities, and results. It consists of a pre-test module that focuses on: education, values, FAQs, patient stories/vignettes and the selection of results. It also includes a post-test check-in module to support users while they wait for results and a function that allows clinicians to upload results for patients to review before or after clinical consults. **Usability testing:** We conducted 25 usability tests with patients, the general public and genetics practitioners (15/25 female; mean age 41 years; 5/25 diagnosed with cancer). Participants were enthusiastic about the application, found it easy to navigate and comprehend. Participants recommended clarifying content and outlining the purpose of tasks. **Acceptance testing:** The application is currently undergoing qualitative and quantitative acceptability testing with a sample of patients and the general public (n=20). The final application will then be evaluated in a RCT with patients undergoing genomic testing. **Conclusions:** We created and tested an interactive, patient-centered application to optimize delivery, access and quality of care for pre- and post-test genomic testing, counselling, and return of results adaptable to any testing platform and setting.

PrgmNr 2520 - Building Genomic Data Science Capacity for Health Discovery and Innovation in Africa

View session detail

Author Block: **T. B. Mersha**[1,2], S. W. Harrar[3], M. Altaye[1], T. Astatkie[4], T. Awoke[5], Z. Shkedy[6], J. Beyene[7]; [1]Cincinnati Children S Hosp. Med. Ctr., Cincinnati, OH, [2]Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [3]2Dept. of Statistics, Univ. of Kentucky, Lexington, KY, [4]Faculty of Agriculture, Dalhousie Univ., Truro, NS, Canada, [5]Dept. of Epidemiology and Biostatistics, Univ. of Gondar, Gondar, Ethiopia, [6]I-BioStat, Hasselt Univ., Diepenbeek, Belgium, [7]McMaster Univ, Hamilton, ON, Canada
Disclosure Block: **T.B. Mersha:** None.
As rapid technological advances make it routine for genomics researchers worldwide to generate increasingly large, complex, and diverse datasets, the challenges of managing and leveraging to efficiently use such data also increase. Addressing these challenges require the collective cutting-edge expertise of computational, statistical, and genomic scientists along with domain knowledge researchers. Genomic data have boundless potential across Africa due to its high levels of genetic diversity (10% more DNA variation than the current reference genome) and disproportionate rate of infectious diseases (25% of the world disease burden). Although methods and applications of data science in genomics have been rapidly evolving over the past couple of decades in resource-rich countries and helped accelerate discovery of new knowledge, the situation is bleak in resource-limited settings such as most countries in Africa primarily due to lack of well-trained genomic data scientists. In this presentation, we describe a roadmap for building the capacities of the next generation of African scientists in genomic data science with the overarching goal of spurring health discovery and innovations. We propose a sustainable infrastructure for data security, privacy, access, analysis and visualization that empower and facilitate large-scale genomic studies, as well as the implementation of genomic medicine. We also outline current and past multiple international initiatives (e.g., *H3Africa,* DSI-Africa, Bridge2AI), potential challenges and future directions for genomics data science in Africa.

PrgmNr 2521 - Challenges to develop TBL modules that integrates genetics and genomics with basic and clinical disciplines in the first year of a Medical School

View session detail

Author Block: **F. T. Lima**, T. A. R. Silva, W. O. Pereira; Faculdade Israelita de CiÃªncia da SaÃºde Albert Einstein, Sao Paulo, Brazil
Disclosure Block: **F.T. Lima:** None.
BACKGROUND: Active methodologies are increasingly essential in the modern teaching process, as well as the delivery of an integrated curriculum, which presupposes approximation of various disciplines into meaningful associations to focus upon broader areas. Team Based Learning (TBL) seems to be a suitable method for such demands. Our Medical School uses TBL as an active methodology in the first 4 years. In two 20-week courses, during the first 2 semesters, Genetics and Genomics are taught in a single discipline that also covers Cell Biology, Molecular Biology, Biochemistry and Biophysics, and many challenges were faced developing TBL modules. AIM: Describe our experience in developing integrated TBL modules and they were approached. METHODS: Teaching plans, syllabus, pre- and in-class materials from 10 semesters (Feb 2016 to Nov 2020) were reviewed, and faculty was interviewed to provide understanding of challenges and how they overcame them. RESULTS: There were 2 TBL modules during first semester and 5-9, on the second, related to Genetics and Genomics. At each semester, until Mar 2020, at least 1 new TBL module per year was developed, totalizing 22 different modules. Development of objectives, content and appropriate application exercises integrating the different subjects, followed by lack of pre-class material in appropriate volume and difficult level were the main challenges. Literature search from the standpoint of disciplines in isolation to later find common ground, exploration of studentâ□□s early clinical experiences, clinical-research partnership and multidisciplinary discussion rounds provided clarification of objectives and ideas for exercises. Feedback from faculty from clinical years also ensured vertical integration between basic and clinical sciences. Creative pre-class videos and materials, using clippings and excerpts of different texts were developed. Reading guidelines are carefully prepared, often in the form of questions, to ensure that the student understands what goals to achieve when reading the material. Additional challenges were faced after Mar 2020, with Coronavirus Pandemic, when we migrated to an online format, the application exercises had to be revised, and several alternatives were tested. The biggest challenge was to ensure the simultaneous response, part of the 4S recommended by the traditional TBL method. CONCLUSION: Developing TBL modules integrating genetics and genomics education with basic and clinical sciences in the first year of a medical school brought many challenges. Careful and creative preparation and team work were essential to overcome them.

PrgmNr 2522 - Consumer-initiated genetic testing: A look at the current landscape in the United States

View session detail

Author Block: **J. Furnival**, S. Bradley, H. Andrighetti, S. D. Crawford; SoundRocket, Ann Arbor, MI
Disclosure Block: **J. Furnival:** None.

Consumer genomics companies offer an array of testing options, including tests for paternity, ancestry, nutrigenomics, carrier status, genetic health risk, and more. The tests vary in quality and completeness; while some tests are clinical grade performed in CAP/CLIA-certified labs, many are not, and laboratory methods range from SNP-based genotyping to whole genome or exome sequencing (WGS/WES). How does a consumer navigate these options to make an informed choice about genetic testing by browsing consumer genomics websites? We set out to understand the current landscape of online consumer-initiated genetic testing options in the U.S., with an emphasis on understanding the scope of health-related genetic testing in 2021.

We aimed to capture all U.S. companies offering genetic tests marketed to consumers. We included both true direct-to-consumer (DTC) genetic testing companies and those that offer products using a consumer-initiated, physician-mediated model. We have assembled a comprehensive, up-to-date list of U.S.-based companies offering consumer-facing genetic testing services. As of June 2021, we have analyzed the websites of 173 companies and wherever available, recorded which test(s) are offered, who initiates the test, who orders the test, type of DNA analysis performed, and lab accreditations or authorizations. To determine who initiates testing, we looked at the language (e.g.: use of â□□your patientâ□□ or â□□youâ□□), tone and user experience of the websites.

We will share a snapshot of the current consumer genomics market along with key findings related to genetic health tests - how many companies are offering physician-mediated genetic health testing, interpretations based on raw data analysis, products based on WES or WGS, or calculation of polygenic risk scores. We will also share the challenges of our approach, including the difficulty of using websites as our data source, and of generating a current list in this high-turnover emergent industry. Our final discussion will include the limitations of only including U.S.-based companies given the international nature of the industry and how online tests can be borderless.

As consumer genomics companies expand their health-related products, more options are emerging that can provide important medical insights. Without a governing body overseeing this industry as a whole, it is difficult to know what is available to consumers, and it is essential that those in the clinical genetics space work to understand what their patients may be accessing.

PrgmNr 2523 - Development and evaluation of digital decision aids with diverse patients: a systematic review

View session detail

Author Block: **S. Krishnapillai**[1,2], E. Adi-Wauran[1,2], Y. Bombard[1,2]; [1]Inst. of Hlth.Policy, Management & Evaluation, Univ. of Toronto, Toronto, ON, Canada, [2]Li Ka Shing Knowledge Inst., St. Michaelâ□□s Hosp., Toronto, ON, Canada
Disclosure Block: **S. Krishnapillai:** None.
Background: There is increasing priority and effort to engage underserved populations in genetics research and testing to mitigate existing health disparities. It is critical to clarify if decision aids (DAs) support diverse communities in their decisions to participate in genetics research and testing. This study will review literature related to design and evaluation of digital DAs with diverse populations. It aims to characterize developments in genetics and glean insights from other fields to collate strategies that support research for this audience and their digital needs. **Methods:** MEDLINE database was searched for quantitative, qualitative, and multi-method studies of DAs with adult populations composed of more than 50% of a diverse group as defined by Flaskerud & Winslowâ□□s conceptual framework of vulnerable populations. All articles were screened first at the title/abstract level and at full-text stage by two reviewers. Type of study, population and DA characteristics, modes of cultural tailoring, and digital considerations were abstracted from included studies. **Results:** 786 articles were identified through database search, ten met inclusion criteria and reported on eight digital DAs of which five examined cancer care, and three focused on birth choices, gender dysphoria or kidney disease. No genetic tools were captured in this search. Nine studies focused on ethnic/racial minority groups and one study on transmen. Studies reported combinations of incidence of disease, mortality, disparities in social determinants of health, and barriers/facilitators to decision-making as the rationale for selecting a population. All eight tools culturally tailored their DAs in the development stage using in-depth interviews or focus groups. To understand participantsâ□□ digital needs, two studies administered surveys to measure digital fluency or self-reported digital efficacy. The majority of participants had high mean scores in computer, e-mail, and web fluency, and self-reported their computer skills highly. **Conclusions:** Our systematic review found no genetic DAs that were designed or supportive for diverse groups, exposing a critical gap in efforts to foster broader recruitment of diverse participants in genetics research and testing. Culturally tailoring interventions with ethnic minorities should take into account language accessibility, involving family and community, and avoid overly simplistic or clinical narratives. Conceptions of diversity need to extend beyond only race/ethnicity to deliver tailored care that involves the breadth of social determinants of health that can influence health disparities including digital literacy.

PrgmNr 2524 - Divergent perceptions of meaning and utility of genetic testing results between families, clinicians, and laboratory directors

View session detail

Author Block: **C. Berrios**, E. Donis De Miranda, K. Pina, E. Hurley, J. Lantos; Childrens Mercy Kansas City, Kansas City, MO
Disclosure Block: **C. Berrios:** Grant/Contracted Research Support (External); Sanofi-Genzyme.
Background: Exome sequencing is frequently used for testing in children with complex symptoms and no obvious diagnosis. We wanted to better understand the responses of all stakeholders (parents, clinicians, and laboratory directors) in cases with varying types of results. Methods: We randomly selected exome sequencing cases from each of three groups of results: those in which the lab reported a diagnostic variant, no diagnostic variant but one or more VUS, and no variants. We recruited parents for in-depth interviews about the reason for testing, meaning and impact of results, and uncertainties or dilemmas resulting from the testing. For each parent interviewed, the ordering physician (OP), genetic counselor (GC) (if involved), and signing laboratory director (LD) were invited to complete interviews on the same topics. Interviews were transcribed, coded, and analyzed using grounded theory-based methodology. Results: Interviews have been analyzed for 3 parents whose child received a diagnostic finding, 3 parents whose child had only VUS, and 1 parent whose child had no reported variants; all ordered by a medical geneticist. OPs for all cases have been interviewed, GCs for 4 cases, and LDs for 5 cases. Parents and providers sometimes had differing views of the meaning and utility of results. Parents valued a diagnostic result that validated their concerns and directly guided clinical care. However, when diagnostic results were not instrumental in facilitating better treatment, parents could feel the result was inconclusive or did not provide helpful information, despite enthusiasm from OPs and LDs about a â□□clear cutâ□□ and useful result. In some cases with VUS, parents relayed extreme disappointment or irrelevance of results. In these cases, OPs and GCs perceived value in ruling out some genetic causes and hope in continued research on variants, though GCs often accurately relayed the parentsâ□□ views. Parents also discussed greater reassurance of a non-genetic etiology from negative results than providers. Differences between parent and provider views did not seem to reflect differences in factual understanding about the variants, but in the meaning and value of the test or result for the child and family. Interviews are ongoing and include purposeful sampling of cases with non-geneticist OPs. Conclusions: Providers should be aware of potential differences in professionalsâ□□ and parentsâ□□ assessment of the utility and limitations of genetic testing, especially when a result does not directly impact patient care. Further research should explore ways to align perceptions about the value of testing and results and assess ethical and psychosocial issues that may arise.

PrgmNr 2525 - Engagement in science: virtual interaction as a flexible tool

View session detail

Author Block: **O. M. Amaral**[1,2]; [1]Inst. Natl. de SaÃºde Ricardo Jorge (INSA), Porto, Porto, Portugal, [2]Univ. of Porto (CECA- ICETA), Porto, Portugal
Disclosure Block: **O.M. Amaral:** None.
The pandemic lockdowns forced a rapid switch to digital platforms. Lockdowns closed schools and laboratories pushing scientist and educators to online activities.
Digital platforms for communication are widely available for free use. These platforms, although different, are intuitive and user friendly, therefore facilitating this sudden imperative shift from in-person to virtual communication.
In an effort to bring services or science information to people new solutions were adopted. Although in-person activities were interrupted, alternative solutions were set up in order to engage wide participation. Such examples can be found among various types of initiatives regarding human genetic disorders which successfully switched to online versions.
After more than a year of digital communication, the outcome regarding interaction between peers; between students and educators; and even between common citizens and scientists can be assessed. With this work specific situations are revisited, indicating points of â□□virtual fatigueâ□□ as well as points of â□□virtual satisfactionâ□□.
Future perspectives on the advantages of using digital platforms will also be presented as part of a project, in the area of Human Genetics, expected to start in the coming year.
Keywords: virtual communication; digital methodologies; Human Genetics education.

PrgmNr 2526 - Free digital genomic educational resources: evaluation of their international value and uptake

View session detail

Author Block: **E. Tobias**[1,2], The University of Glasgow (UK) Medical Genetics and Genomics Education Team; [1]Univ. of Glasgow, Glasgow, United Kingdom, [2]West of Scotland Ctr. for Genomic Med., Queen Elizabeth Univ. Hosp., Glasgow, United Kingdom
Disclosure Block: **E. Tobias:** None.
BACKGROUND: The need for access to reliable digital educational resources for genetics and genomics is increasing significantly. This is partly due to the rapid growth in access to genetic and genomic testing by not only genetics health professionals but also, increasingly, by non-genetics specialists and the public. Understandably, the ASHG and other international societies are prioritising genetic and genomic education. Here, the creation of multiple free digital genetic and genomic educational resources, and the lessons learned in the process, will be discussed. **METHOD:** Members of an award-winning UK Glasgow University genetics and genomics teaching team have created a free educational range of resources, including: five smartphone apps (covering genomics and bioinformatics terminology & inheritance mechanisms), two free recently-launched massive open online part-time courses (MOOCs) and two websites. All are suitable for genetics specialists plus non-genetics professionals and students. One of the websites (www.EuroGEMS.org), for multiple audiences, was created on behalf of the European Society of Human Genetics (ESHG), as a guide to >110 free high-quality online worldwide resources (many USA-based). It was recently fully translated into Spanish by genetics professionals. **RESULTS:** The five free smartphone genomics apps (www.genomicsapps.org) have been installed by >5000 people in >70 countries and rated 4.9 stars. The new MOOC on Medical Genomics (https://www.futurelearn.com/courses/harnessing-the-power-of-genomics-in-medicine) covers variant filtering, naming, classification, sequencing methods, databases and precision medicine. It has, in 4 months, gained >1200 learners from >110 countries (particularly UK, India & USA), with a peak age range of 26-35. Learners contributed >650 comments and have awarded the course a 5-star rating, commending its clarity; explanations of complex topics, terminology and databases; and its online interactivity. The www.EuroGEMS.org website has been visited from 120 countries, 21% from outside Europe. Users from the USA make up a significant proportion of all of the resourcesâ□□ users but the data suggest that most potential users remain unaware of the resourcesâ□□ existence, possibly due to search engine algorithms. **CONCLUSIONS:** There is growing worldwide use and appreciation of these various digital resources, many of which can be helpful in supporting and supplementing in-person face-to-face teaching. Challenges include the time required for the resourcesâ□□ creation and the difficulties in making potential beneficiaries in the USA and elsewhere aware of their existence.

PrgmNr 2527 - Including children in a polygenic risk score return of results pragmatic trial: Considerations and decisions by consensus in the eMERGE Network

View session detail

Author Block: **C. Prows**[1], S. Wood[2], E. Bhoj[3], J. J. Connolly[3], S. L. Van Driest[4], N. S. Abul-Husn[5], W. K. Chung[6], E. W. Clayton[4], J. Gundelach[7], I. A. Holm[8], E. M. McNally[9], M. F. Myers[1], M. Sabatello[6], eMERGE Network; [1]Children's Hosp. Med. Ctr., Cincinnati, OH, [2]Univ. of Cincinnati Academic Hlth.Ctr., Cincinnati, OH, [3]Children's Hosp. of Philadelphia, Philadelphia, PA, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]Columbia Univ., New York, NY, [7]Mayo Clinic, Rochester, MN, [8]Boston Children's Hosp., Boston, MA, [9]Northwestern Univ., Chicago, IL
Disclosure Block: **C. Prows:** None.
Polygenic risk scores (PRSs) to assess risk for common, complex diseases with adult onset such as breast cancer and coronary heart disease have become commercially available and are used in some clinical settings. Published PRSs for childhood onset diseases such as asthma and type 1 diabetes are comparatively more recent and not used in clinical settings. Guidance on PRS return to pediatric research participants is lacking. The electronic MEdical Records and Genomics Phase IV (eMERGE IV) Network consists of ten clinical sites, two of which are pediatric medical centers. Each site is expected to enroll 2500 children and adults into a pragmatic trial in which PRSs will be calculated for a number of conditions and incorporated along with family history and clinical factors into a Genomic Informed Risk Assessment (GIRA) and returned to participants and their electronic health records (EHRs). The Network considered over 60 different conditions for PRS implementation and used a consensus process to select 17 conditions for PRS validation and potential inclusion into a comprehensive GIRA. A pediatric sub-workgroup developed recommendations for PRS appropriate for testing and return to pediatric participants. Criteria appraised included scientific (e.g. sensitivity/specificity, existing validated PRS), ethical (e.g. actionability in childhood, risk vs. benefit of intervention), and practical (e.g. local expertise, age range for study recruitment, EHR access for children and parent(s)) considerations. Health promotion, disease prevention and early treatment interventions at age ranges 0 - 2, 3 - 12, and 13 - 17 years were assessed across potential pediatric conditions and used as a gauge for site votes on age inclusion criteria for the pragmatic trial. Thirteen conditions were removed from consideration due to the lack of clinical actionability during childhood or Network early determination that a PRS did not meet criteria for implementation. Rigorous PRS validation processes and comparisons followed for four conditions deemed actionable during childhood: asthma, obesity, and types 1 and 2 diabetes. Key decisions and rationales as well as remaining ethical issues to be studied will be shared. Our experience will inform others planning to calculate and return childrenâ□□s PRS in research and clinical settings.

PrgmNr 2528 - Integration of Genetic Counseling Services into an Immunodeficiency Clinic: Roles of the Genetic Counselor and Impact on Patient Evaluation and Care

View session detail

Author Block: **E. A. Varga**, P. Mustillo, B. T. Prince, R. S. Abraham; Nationwide Children's Hosp., Columbus, OH
Disclosure Block: **E.A. Varga:** None.
Background: Next generation sequencing is increasingly applied to diagnose inborn errors of immunity. Barriers to integration of genetic testing into clinical immunology practice may include the need to evaluate rapidly evolving genomic technologies without subspecialty training, unfamiliarity with the informed consent processes for genetic testing, difficulties in navigating insurance reimbursement and concerns related to variant interpretation, secondary findings and medicolegal responsibilities associated with patient and family follow-up. **Objective:** To evaluate uptake of genetic counseling services offered through an immunodeficiency clinic at a tertiary referral pediatric hospital over the course of one year and assess implications for patient care and outcomes.
Methods: The number or referral requests from immunologists, genetic counselor roles, genetic test results and outcomes were recorded. **Results:** Forty-six patients were referred for genetic counseling by a clinical immunologist. Three of these patients had a previous molecular diagnosis where follow-up genetic counseling was requested. Four patients had previous molecular testing where review of past test results and reinterpretation led to revised counseling or diagnosis. Six patients received a molecular diagnosis after further genetic testing. Five patients required genetic counseling related to incidental findings/carrier status. In addition, the genetic counselor contributed to the care team by helping to evaluate and discuss genetic testing options with clinicians and patients, assist with insurance prior authorization for genetic testing, obtain informed consent for genetic testing, enroll patients in genomic research protocols and identify patients appropriate for referral to a medical geneticist. Counseling-centric roles included pedigree intake and risk assessment, discussion of inheritance, recurrence risk and/or family planning options with patients and their family members. Genetic testing was coordinated in sixteen relatives, and cord-blood banking was facilitated in one case. A molecular diagnosis altered treatment in two cases by prompting hematopoietic cell transplant (HCT). **Conclusion:** Immunologists are likely to benefit from access to a genetic counselor to facilitate genetic testing, provide input on interpretation of genetic test results, and to ensure appropriate follow-up family testing, education, support and coordination of care.

PrgmNr 2529 - International education, awareness, celebration, and commemoration around the bicentennial of Gregor Mendelâ□□s birth

View session detail

Author Block: **J. J. Mulvihill**[1,2], J. Carda[3], M. Macek[4], M. Genuardi[5,6], C. Easter[7], D. J. Fairbanks[8], M. S. Lindee[9], J. SekerÃik[10], V. Adam[11], O. DostÃil[12], B. KÅ□ÃÅ¾ovÃi[13], Å . PospÃÅiilovÃi[12]; [1]Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK, [2]Natl. Human Genome Res. Inst., Bethesda, MD, [3]Agency Spolecne, Brno, Czech Republic, [4]Motol Univ. Hosp.-Charles Univ. 2 Faculty of Med., Prague, Prague, Czech Republic, [5]Fondazione Policlinico Univ.rio A. Gemelli IRCCS, Rome, Italy, [6]UniversitÃ Cattolica del Sacro Cuore, Rome, Italy, [7]Natl. Museum of Natural History, Washington, DC, [8]Utah Valley Univ., Orem, UT, [9]Univ. of Pennsylvania, Philadelphia, PA, [10]Moravian Museum, Brno, Czech Republic, [11]Mendel Univeristy, Brno, Czech Republic, [12]Masaryk Univ., Brno, Czech Republic, [13]Mendel Museum of Masaryk Univ., Brno, Czech Republic

Disclosure Block: **J.J. Mulvihill:** None.

THE PERSON: Johann Mendel was born July 22, 1822, and died as Abbot Gregor Mendel on January 6, 1884. After 8 years of experiments and 18 months of compiling and interpreting his results, he founded the science of genetics, giving two lectures at the Natural Science Society in Brno (Czechia) and publishing two papers on inheritance in the societyâ□□s journal. His achievements hold lessons for today for his abilities to apply quantitative tools from physics and mathematics to qualitative biologic data, to anticipate the mechanisms of Darwinian evolution, and to improve horticulture and animal breeding, while being a faithful practicing priest and leader of his religious community. He can be held as an exemplar of mentoring (in both directions, giving and taking), thriving at the interdisciplinary interactions of botany, evolution, mathematics, physics, and agriculture, contributing to improved fruit trees, ornamental flowers, horticulture, meteorology, and beekeeping, and being a scientist-citizen, a community and social activist against inqualities, bank director and president, and a favorite member of his biologic and monastic families. He illustrates the parallel and compatible realms of science and religion and the pros and cons of solitary science and team science. His data, an early example of big data science, were challenged as too good to be true, a trope best handled with reliable research and perhaps some humor. **METHODS:** To mark the occasion, planning has involved virtual meetings with an international group representing diverse interests and perspectives. With the goals of Improving science literacy and awareness, education, appreciation of science, celebration, inspiration for trainees, students, and the public, international collaboration, and enjoyment, the vision is to increase global awareness among geneticists, other scientists, and the public of the current meanings of Mendelâ□□s life and work by helping to implement international programming designed to educate, celebrate, commemorate, and bring awareness and equity to diverse audiences. **RESULTS TO DATE:** Release of new monograph (G Mendel: *Versuche Ã¼ber Pflanzen-Hybriden, Experiments on Plant Hybrids: New Translation with Commentary.* S MÃ¼ller-Will, K Hall, O DostÃil, Eds., Brno, Masaryk University Press, 2020), improved exhibits at some museums, including the Smithsonian, an international competition for new statue of Mendel, restoration of Mendelâ□□s greenhouse, launching a bicentennial website https://www.mendel22.cz/, and announcement of the Mendel Festival and Genetics Conference, July 20-23, 2022. In February 2022, AAAS will have a Mendel session. Your input is welcomed!

PrgmNr 2530 - Participant Engagement and Cancer Genome Sequencing (PE-CGS) Network

View session detail

Author Block: **L. E. Mechanic**[1], J. C. Tiner[1], W-Y. Chou[2], K. K. Filipski[1], E. M. Gillanders[1]; [1]Natl. Cancer Inst./DCCPS/EGRP, Bethesda, MD, [2]Natl. Cancer Inst./DCCPS/BRP, Bethesda, MD
Disclosure Block: **L.E. Mechanic:** None.
While recent genome characterization efforts have been very successful, research gaps remain especially for minority and underserved populations. The Cancer Genome Atlas (TCGA) program has analyzed over 20,000 samples for 33 different cancer types, yet only an estimated 15% of TCGA cases are from non-Caucasian participants and only 3% of from Hispanic or Latino participants. One method of addressing these knowledge gaps and increasing diversity in genetic research is direct participant engagement. Direct participant engagement may include interacting with participants via the web, social media, or in collaborations with patient groups or organizations. As a part of the Cancer Moonshot^{SM} Initiative, the National Cancer Institute (NCI) has launched the Participant Engagement and Cancer Genome Sequencing (PE-CGS) Network which falls under the Blue Ribbon Panel (BRP) Recommendation A: Network for Direct Patient Engagement. In the recommendation, the BRP called for NCI to â□□enlist direct patient engagement through a federated network where patients will be offered comprehensive tumor profilingâ□□. The panel further noted the importance of directly engaging with patients to facilitate participation in research and ensure patients are respected and have access to the research enterprise. Moreover, the panel stated that efforts to reach minority and underserved populations should be a high priority. The PE-CGS Network includes several U2C Research Centers and one U24 Coordinating Center. The overall purpose of the PE-CGS Network is twofold: 1) To promote and support direct engagement of cancer patients and post-treatment cancer survivors as participants in cancer research; and 2) To use such approaches for rigorous cancer genome sequencing programs addressing important knowledge gaps in the genomic characterizations of tumors in areas such as, but not limited to: Rare cancers or rare cancer subsets; highly lethal cancers; cancers with an early age of onset; cancers with high disparities in incidence and/or mortality; or cancers in understudied populations. The PE-CGS Network aims to function as a collaborative network allowing PE-CGS U2C Research Centers to address common issues, share best practices and lessons learned, and utilize common methods where appropriate. Through this network, NCI aims to address research gaps in the genomic profiles of cancer, determine effectiveness of a direct participant engagement approach, generate a shared resource for the scientific community, and provide insights into the development and sustainability of a larger network for direct participant engagement.

PrgmNr 2531 - Public Health Genetics Week 2021: A Public Awareness Campaign

View session detail

Author Block: **M. Lyon**; American Coll. of Med. Genetics and Genomics, Bethesda, MD
Disclosure Block: **M. Lyon:** None.
Purpose Public health genetics plays a vital role in both genetics and public health, the field is often overlooked within the larger context of either specialty. To raise awareness, and celebrate the efforts of the system, the second annual Public Health Genetics Week was held from May 24-28,2021.
Methods The National Coordinating Center for the Regional Genetics Networks (NCC) held the second annual Public Health Genetics Week. In collaboration with federal agencies, state agencies, regional organizations, and non-profits, NCC developed materials and held events for Public Health Genetics Week. The week consisted of daily themes (â□□What is Public Health Genetics?â□□, â□□Who is Involved in Public Health Genetics?â□□, â□□What are Public Health Genetics Programs?â□□ â□□Public Health Screeningâ□□, and â□□Public Health Genetics Resourcesâ□□) and included public health genetics events (nightly screenings of *Ken Burns Presents The Gene: An Intimate History*, Reddit AMA, Public Health Genetics Student Panel, Digital Escape Room, Facebook Live, and Twitter Chats). The daily themes and events targeted stakeholders of the public health genetics system, as well as the general public, to raise awareness about public health genetics. **Results** Across Twitter, Facebook, Instagram from May 24-28, there were 299 mentions of Public Health Genetics Week, utilizing either #PublicHealthGenetics or #PHGW hashtags. On TikTok, videos were viewed more than 2800 times. Additionally, the conversation reached over 524,000 individuals across Facebook, Instagram, and Twiter. Throughout the week, individuals and families, federal agencies, organizations (for-profit and non-profit), engaged in conversations about public health genetics and shared resources that can be utilized by the system. *This project is supported by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS) under Cooperative Agreement #UH9MC30770-01-00 from 6/2017-5/2020 for $800,000 per award year. This information or content and conclusions are those of the author and should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS or the U.S. Government.*

PrgmNr 2532 - RNA-Seq and recurrence risk testing for breast cancer in the context of an undergraduate course: Implications for patient education

View session detail

Author Block: **P. Soneral**; Bethel Univ., Saint Paul, MN
Disclosure Block: **P. Soneral:** None.
Rapid advances in RNAÂseq technology and bioinformatic analysis of gene expression has normalized the use of precision diagnostics and recurrence risk testing for breast cancer patients. While these diagnostic approaches have numerous benefits, patients often lack the scientific literacy to evaluate data models - and their uncertainties - for enhancing their health decision-making. To increase access to scientific knowledge for stakeholders, we designed an educational module that evaluates the validity of RNA-seq gene expression as a tool for subtyping and determining the prognostic signature of a human breast tumor. Using a crowdsourced approach in the context of an undergraduate course, we recreated the workflow of prominent precision oncology companies - we prepared a whole transcriptome library from a single breast cancer patient, and used multiplex sequencing on twelve uniquely barcoded Illumina preparations. Datasets obtained from the sequencing were subjected to quality control analysis using the Green Line of DNA Subway (FAST QC, FAST X, Kallisto, Sleuth algorithms). LogÂ transformed gene expression ratios were used to determine the molecular subtype of the tumor. To calculate the recurrence risk, the Oncotype DX model was used, including 16 cancer genes and 5 reference genes. Results from the proof-of-concept experiment indicated significant overexpression of ER and PR genes (p25, suggesting high risk for 10 year distant recurrence (95% CI). Further exploration of the model indicated that normalization to aberrantly expressed reference genes can greatly skew the recurrence risk calculation, highlighting limits to the predictive power of this tool. Taken together, our findings illustrate the validity of RNA-seq based tumor subtyping through an educational module, and underscores the importance of equitable access to genomics education as a form of patient self-empowerment and decision-making. We discuss the future implications of this work for education and policy.

PrgmNr 2533 - The role of general practitioner in management of rare diseases

View session detail

Author Block: **P-H. Roux-Levy**[1], Y. Perrard[1], K. Mazalovic[1], C. Binquet[2], N. Meunier-Beillard[2], L. Faivre[3], C. Lejeune[2]; [1]Dept. of Gen. Practice, Dijon, France, [2]Inserm, CIC1432, Clinical Epidemiology Unit, Dijon, France, [3]Hosp d' Enfants, Dijon, France
Disclosure Block: **P. Roux-Levy:** None.
Background: The management of patients with rare diseases (RD) presents many challenges including diagnosis, care coordination and orientation in the health system. The general practitioner (GP) is the dedicated actor who provides long-term care to the patients and their families. The aim of this study was to determine the roles of the GP in management of the patient with RD.
Methods: We used a qualitative study by monographs. From March to October 2020, semi-structured interviews were conducted by telephone with families of adult patients who have RD with intellectual disability and with the health professionals who took part of their management. Patients were recruited through reference centers, support group or GP to have very different type of management. The interviews were transcribed and analysed by two independant investigators. A grounded theory-based analysis was realised.
Results: Eight monographs were conducted with 20 participants. The onset of the first symptoms and the disability will mark a break point with the idealized child and family. During the life of these patients, the parents, and more often the mother, are subjected to three challenges in face of which they will have to position themselves in various ways: the family balance, the caregiving and the care coordination. The parents' positioning regarding these issues will vary according to the availability of psychosocial, technical, organizational and administrative resources, but also according to their experience with their child's illness and disability. They will invest each of these positions in a more or less intense way and thus build their new identity. The GP will be involved in different ways depending on this construction.
Discussion: Although the GP has a vocation as well as all the resources to ensure the care coordination, he must take into consideration the identity construction process of the natural caregiver around the illness and the disability of his child. If the caregiver constructs his or her identity through the care coordination, the GP must position himself as a support by helping the parent to coordinate care and by providing psychological support. Once this process has evolved, he will be able to take over the care coordination and thus reduce the parent's burden.

PrgmNr 2534 - Transitioning to a virtual training model for building capacity in genomic analysis skills across world regions

View session detail

Author Block: **A. Matimba**[1], M. Aslett[1], I. Malta[1], S. Aron[2], D. Hughes[1]; [1]Wellcome Connecting Sci., Cambridge, United Kingdom, [2]Univ. of Witwatersrand, Johannesburg, South Africa
Disclosure Block: **A. Matimba:** None.
Background: Short courses offer an effective way for scientists to acquire specialised skills relevant to their work, contributing to addressing capacity gaps for genomics and bioinformatics skills, particularly in regions where the needs continue to escalate. Following a needs assessment survey among scientists in Africa, we identified genome-wide association studies (GWAS) as a priority topic for training. The COVID-19 pandemic then necessitated development of online learning formats leading to our first virtual course delivery. The â□□Human Genomic Epidemiology in African Populationsâ□□ course, aimed to provide genomic scientists with â□□hands-onâ□□ practical exercises on the study design, data processing, analysis and interpretation to understand the genetic architecture of complex human traits/diseases. This was complementary to genomics capacity building initiatives in Africa and facilitated exploration of design and development of virtual training across world regions.
Methods: The week-long course was developed collaboratively with 12 human genomics and epidemiology experts. Applicants were selected based on potential benefits to their work or research immediately or in the short term. In addition to online resources, software and relevant datasets were compiled onto a virtual machine, which was disseminated to participants providing all analysis tools and resources in one place. Preparation of participants emphasised a need for minimum technical specifications and internet connectivity. A combination of virtual teaching strategies included watching lectures online, live discussions and hands-on practical analysis exercises in small tutorial groups. The course used online meeting platforms, text chat resources and a web-based repository for training materials.
Results and conclusions: The course was delivered to 22 participants from 8 countries in Africa. Experiencing the delivery method for the first time, participants reacted positively to the effectiveness of the virtual platforms for delivery of a highly hands-on practical advanced genomics course. Over 90% of participants indicated that the course met their expectations, was relevant for their work and achieved desired learning outcomes. A follow-up survey will determine the impact of the course on their work, institutions and networks. Modelled in a similar way, and taking lessons learnt, this course will be tailored for virtual delivery in other regions. In conclusion, we identified strengths and opportunities for further refinement and tailoring of virtual training approaches for bioinformatics to audiences based in Africa, Asia and Latin America.

PrgmNr 2535 - User-centered design of genetic test reports leads to improvements on self-reported and objective measures

View session detail

Author Block: **G. Recchia**[1], A. Lawrence[1], L. Freegard[2], L. Capacchione[3], G. Chandratillake[4,5], A. L. J. Freeman[1]; [1]Univ. of Cambridge, Cambridge, United Kingdom, [2]Severn Deanery, Natl. Hlth.Service, Bristol, United Kingdom, [3]East and North Hertfordshire NHS Trust, Hertfordshire, United Kingdom, [4]Univ. of Cambridge Inst. of Continuing Ed., Cambridge, United Kingdom, [5]East of England Genomic Med. Ctr., Cambridge, United Kingdom
Disclosure Block: **G. Recchia:** None.
As genetic testing becomes more widely accessible and cost effective, there is an increasing need for non-specialist healthcare professionals to order, interpret and action a variety of genetic reports. Previous research indicates that patients and clinicians frequently have difficulties interpreting existing reports, particularly for *BRCA*. We have previously published general report recommendations and specific findings for cystic fibrosis carrier testing based on a user-centered design research project. Here we report the results of follow-up studies employing a total of six rounds of interviews and analysis to redesign report templates for *HFE* and *BRCA1*. These involved interviews with 25 healthcare providers and 41 members of the public, including but not exclusive to individuals having prior experience with *HFE* or *BRCA* testing. Feedback was gathered on wording, clarity of the next steps that should be taken (actionability), and other dimensions, and reports were revised accordingly. Report design was followed by preregistered evaluations comparing report templates having this novel design with corresponding report templates used within the UK national testing network in randomized factorial (design x template) between-participants designs. The sample consisted of (blinded) members of the public recruited by online access panels, n = 376 (*HFE*) and 456 (*BRCA1*). Endpoints included objective and subjective comprehension, communication efficacy, actionability, worry, and explainability (â□□I would be able to explain what the result means to people in my family who may need to knowâ□□). The modified templates scored higher on the communication efficacy scale, and were rated easier to understand, easier to explain and more actionable (each test *p* p

PrgmNr 2536 - Efficient delivery of *FMR1* across the blood brain barrier using AAVphp construct - feasibility of gene therapy for fragile X syndrome

View session detail

Author Block: **M. T. Velinov**[1], C-L. Dai[2], G. Merz[2], T. Wise[2], J. Goodman[2], H. Meeker[2], T. Adayev[2], N. Dolzhanskaya[2]; [1]Rutgers Robert Wood Johnson Med. Sch., New Brunswick, NJ, [2]NYS Inst. for Basic Res. in Dev.al Disabilities, Staten Island, NY

Disclosure Block: **M.T. Velinov:** None.

Background: Fragile X syndrome (FXS) is the most common single gene disorder associated with intellectual disability and autism. It results from the silencing of gene *FMR1.* The protein product of *FMR1,* FMRP binds multiple mRNA molecules and is a major regulator of protein translation and function. Patients with partially preserved *FMR1* activity due to methylation mosaicism show increased cognitive functioning suggesting that partial supplementation of FMRP may mitigate disease manifestations. Clinical trials using small molecule medications failed to demonstrate significant control of the disease core symptoms. Direct intraventricular administration of *FMR1* expression constructs resulted in uneven distribution of the construct in different brain regions. AAVphp vector constructs were recently shown to provide efficient gene delivery across the blood brain barrier. We therefore aimed to determine the efficiency of *FMR1* -AAVphp constructs delivery in the brain after peripheral administration. Methods and Results: We first used two AAV vectors that harbor fluorescent marker GFP and the human synapsin promoter (scAAV9-hSyn-GFP and scAAVphp.eb-hSyn-GFP, VirovekInc) to assess the efficiency of brain delivery after peripheral administration. We injected $2X10^{13}$ vg/kg in a final volume 200Âµl of sterile PBS in the mouse tail vein in control animals. We observed robust marker expression in neurons. Further, we administered constructs including the *FMR1* gene with both human and mouse synapsin promoter (AAVphp.ebâ□□hSyn-mFMR1IOS7, AAVphp.ebâ□□mSyn-mFMR1IOS7, Virovek Inc) in the amount $2X10^{13}$ vg/kg in a final volume 200Âµl of sterile PBS and a control empty AAVphp.eb vector in the peripheral tail veins in 5-month old control mice (FVB.129P2-Pde6b+tyr-ch Ant/J) and in a *FMR1* knockout mouse model (FVB.129P2-Pde6b+tyrc-ch Fmr1tmlCgr/J). Following 4-week post-injection interval, whole brain homogenates were examined for FMRP expression using quantitative FMRP (qFMRPm) method developed for mouse tissue analysis. The assay utilizes the FMRP-specific mAb5C2 to capture the rabbit polyclonal antibody R477 for detection, and an abbreviated recombinant protein, GST-mR7 as standard. In brains of knock out mouse models after construct injections we observed FMRP levels similar to the observed in control animals. No FMRP was noted in knock out models injected with control empty vector. Conclusions: We demonstrated robust *FMR1* delivery and expression in the brain of mouse *FMR1* knock out models after peripheral administration using AAV.php vectors. Our results suggest that further studies using the peripheral AAVphp.eb vector for *FMR1* delivery are warranted.

PrgmNr 2537 - Exon specific U1 snRNA therapeutic strategy to rescue retinal degeneration in familial dysautonomia

View session detail

Author Block: **A. Chekuri**[1], E. Morini[2], E. Logan[2], A. Krauson[1], M. Salani[1], E. Kirchner[1], J. Bolduc[1], G. Romano[3], F. Riccardi[4], F. Pagani[3], L. H. Vandenberghe[5], S. A. Slaugenhaupt[6]; [1]Massachusetts Gen. Hosp./Harvard Med. Sch., Boston, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Intl. Ctr. for Genetic Engineering and Biotechnology, Trieste, Italy, [4]Intl. Ctr. for Genetic Enegnineering and Biotechnology, Trieste, Italy, [5]Harvard Med Sch, Boston, MA, [6]Massachusetts Gen. Hosp. Res. Inst., Boston, MA
Disclosure Block: **A. Chekuri:** None.
Familial dysautonomia (FD) is an autosomal recessive neurodegenerative disorder caused by a splice mutation in the gene encoding Elongator complex protein 1 (*ELP1*, also known as *IKBKAP*). A T-to-C base change in the 5â□□ splice site of *ELP1* exon 20 results in exon 20 skipping with tissue specific reduction of ELP1 protein predominantly in central and peripheral nervous system. In addition to complex neurological phenotype, FD patients also exhibit progressive retinal degeneration severely affecting their quality of life. To test novel splicing-targeted therapeutic approaches, we developed a phenotypic mouse model of FD, *TgFD9; Ikbkap*$^{\hat{I}\square 20/flox}$ which exhibits most of clinical features of the disease while displaying the same tissue specific mis-splicing observed in patients.Here, we report a thorough characterization of the retinae of our FD mouse using SD-OCT and immunohistochemical assays during disease progression. Our findings showed a significant decrease in the thickness of the retinal nerve fiber layer (RNFL) and the ganglion cell layer (GCL) starting from 3 months of age. Retinal whole-mount analysis showed reduction of RGC cell counts from 6 months of age. Histopathological analysis of the optic nerve from FD mice using neurofilament (NF) staining indicated diffuse degeneration of axonic bundles demonstrating that our mouse model correctly recapitulates the retinal degeneration observed in patients. To restore correct *ELP1* splicing defect and rescue retinal degeneration, we have designed a novel splice targeted therapy using modified version of the spliceosomal U1 snRNAs (ExSpeU1s) that permit targeted binding to intronic sequences downstream of the mutant 5' splice site enabling efficient recruitment of spliceosomal machinery. We have analyzed the efficiency of splicing correction in the retina of FD mouse through intravitreal injection of adeno associated vectors (AAV) expressing ExSpeU1.Our findings suggest that our novel FD mouse model exhibit most of the retinal degeneration pathology observed in FD patients. Our in vivo preliminary data demonstrate the valuable therapeutic potential of ExSpeU1 RNA delivery to treat retinal degeneration in FD.

PrgmNr 2538 - Nanoparticle-based non-viral CRISPR/Cas9 delivery platform for targeting thoracic aortic aneurysm associated heterozygous mutations

View session detail

Author Block: **L. Ellis-Aguilar**[1], C. A. Velandia-Piedrahita[2], J. Cifuentes[1], J. C. Cruz[1], C. MuÃ±oz-Camargo[1], R. Cabrera[1]; [1]Univ. de los Andes, BogotÃi, Colombia, [2]FundaciÃ³n Cardioinfantil - Inst. de CardiologÃa - Molecular Biology and High Complexity Tests Lab., BogotÃi, Colombia
Disclosure Block: **L. Ellis-Aguilar:** None.
From cell culture improvement to design of therapies for rare diseases, genome editing using CRISPR/Cas9 has become one of the most attractive tools for researchers due to its versatility, high efficiency and the possibility of reduced off-target effects. However, its clinical application has been hindered by the scarcity of safe and efficient delivery systems. Here, we developed a magnetite-based bionanoconjugate to efficiently deliver, both recombinant Cas9 and single guide RNA (sgRNA). Specifically, we synthesized Magnetite Nanoparticles (MNPs) and functionalized them by the conjugation of a polymer spacer (PEG) and the cell-penetrating peptide Buforin II (BUFII) for superior membrane translocation. Also, we co-immobilized recombinant Cas9 to form MNP-PEG-BUFII-Cas9 nanobioconjugates. We characterized them by Fourier transform infrared spectroscopy (FTIR), thermogravimetric analysis (TGA), dynamic light scattering (DLS) and transmission electron microscopy (TEM). We also carried out biocompatibility analyses including lactate dehydrogenase (LDH) cytotoxicity assays, platelet aggregation assays, hemolytic activity assays and the Ames test for mutagenicity. We evaluated cell internalization and compared it with electroporation. For the Cas9 activity proof-of-concept we first incubated our nanoconjugate with PCR products to evaluate *in-vitro* digestion. Second, we amplified genomic DNA from 4 patients with hereditary aortic aneurysms and designed gRNA for allele specific digesting targeting the genes *PKD1, TGFBR2, COL3A1* and *COL15A1*. Then, we performed Digenome-Seq for off-target effect detection. Third, we delivered the MNP-PEG-BUFII-Cas9 nanobioconjugate to Cos7 and HFF cells to introduce a DNA lesion at *COL3A1.* Finally, we evaluated cell internalization and endosomal escape of the nanobioconjugate by co-localization analysis of confocal images. Overall, we demonstrate the efficacy of MNPs as safe and promising non-viral delivery vehicles for CRISPR/Cas9 localized gene editing to treat heterozygous mutations.

PrgmNr 2539 - Additional data obtained from Exome/Genome Sequencing : two national studies to discuss the risk-benefit balance for implementation in France

View session detail

Author Block: **L. Faivre**[1], F. Houdayer[2], B. GÃ©rard[3], A. Chassagne[4], P. Aurore[5], D. Salvi[5], M-L. Asensio[6], S. Staraci[7], A. Cadenes[8], G. Lesca[2], B. Keren[2], M. Spentchian[2], E. E. GAUTIER[9], A. Baurand[1], C. Sawka[1], G. Bertolone[1], C. Thauvin-Robinet[10], C. Philippe[10], D. HÃ©ron[11], C. Peyron[5], C. Lejeune[12], H. Dollfus[13], C. Binquet[12]; [1]Ctr. de RÃ©fÃ©rence Anomalies du DÃ©veloppement et Syndromes Malformatifs Est, CHU Dijon-Bourgogne, Dijon, France, [2]Service de gÃ©nÃ©tique clinique, Ctr. de RÃ©fÃ©rence Anomalies du DÃ©veloppement et Syndromes Malformatifs Sud Est, Hospices Civils de Lyon, Bron, France, [3]Laboratoire de Biologie MolÃ©culaire, CHRU Strasbourg, Strasbourg, France, [4]Laboratoire de Sociologie et dâ□□Anthropologie (LaSA, EA3189), UniversitÃ© de Bourgogne, Franche-ComtÃ©, Besancon, France, [5]Laboratoire dâ□□Ã□conomie de Dijon (LEDI), EA7467, UniversitÃ© de Bourgogne, Franche ComtÃ©, Dijon, France, [6]INSERM CIC-EC 1432, Dijon Univ. Hosp., Dijon, France, [7]Service de gÃ©nÃ©tique, Ctr. de RÃ©fÃ©rence pour les Maladies cardiaques hÃ©rÃ©ditaires, GH APHP, Paris, Paris, France, [8]Service de gÃ©nÃ©tique clinique, Ctr. de RÃ©fÃ©rence Anomalies du DÃ©veloppement et Syndromes Malformatifs Sud Est, Hospices Civils de Lyon, Lyon, France, [9]CHU DIJON, DIJON, France, [10]FHU TRANSLAD et Ã©quipe GAD INSERM UMR 1231, Franche-ComtÃ©, Dijon, France, [11]Ctr. de RÃ©fÃ©rence Anomalies du DÃ©veloppement et Syndromes Malformatifs, GH APHP, Paris, France, [12]CIC, Inserm 1431, CHRU BesanÃ§on, Besancon, France, [13]Laboratoire de Biologie MolÃ©culaire, CHU Strasbourg, Strasbourg, France

Disclosure Block: **L. Faivre:** None.

With the development of next generation sequencing, additional information (incidental data or secondary data (SD)) may become available. This information, unrelated to the symptoms that justified the prescription of the test, may be of potential interest to patients/families, but may also of additional concern. Some foreign learned societies recommend that the patient be offered a systematic analysis of a pre-established list of so-called 'actionable' genes, while others do not recommend this analysis in the absence of clear arguments about the benefit-risk ratio. Research studies are encouraged. Two complementary French studies are being proposed.

FIND (330 patients, 3 centers, exome sequencing, developmental abnormalities) is analyzing the medical benefits, economic and psychological issues of reporting SD (group 1 late onset actionable diseases, group 2 genetic counselling, group 3 pharmacogenetics). Information about SD were given to participants after a dedicated consultation with a genetic counsellor. Semi-structured individual interviews have been conducted in parents of patients with one SD just after the results, and at 6 and 12 months. 80% of the parents accepted the search for SD in their child, and we showed an influence of the medical discourse on the parentsâ□□ choices. All the DS results of FIND have been given back to parents, and the follow-up interviews up to one year did not show regrets, but it was not without any psychological impacts in group 1. With time, some participants have forgotten the SD results, especially from group 2 and 3, and sometimes, there was confusion with the primary data.

The DEFIDIAG-DS (1200 patients, genome sequencing) will allow access to ACMG SD, not to index cases with ID, but to their parents. Quantitative data, change of opinion/ refusal questionnaire, semi-directive interviews in case of ACMG SD will be proposed at the time of results and at one year. The support of the study allows a deployment in 12 centres instead of 3, which will allow to deepen the impact of the medical discourses, and of the decisions of the parents when it concerns their health.

The results of both studies will increase knowledge on the perception and the impact of SD, and give avenues regarding information, announcement and management.

PrgmNr 2540 - Enhanced diagnosis mapping and exome sequencing identify novel variants for rare diseases in the UK Biobank

View session detail

Author Block: **M. Patrick**[1], R. Bardhi[1], W. Zhou[2], J. E. Gudjonsson[1], L. C. Tsoi[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Massachusetts general Hosp., Broad Inst., Boston, MA
Disclosure Block: **M. Patrick:** None.
While many people are affected by rare diseases, the ability to accurately identify their genetic causes can be limited due to the small sample sizes available. Population biobanks provide extensive genetic and phenotypic information on a large number of individuals, however, we previously showed that rare disease mapping can be suboptimal. We hypothesize that, by applying enhanced disease mapping in a large population biobank with whole exome sequencing (WES) data, we can enhance the identification of associated variants. We used WES data from 167,246 Caucasian individuals (cases and controls) for genetic association of 162 rare diseases (indicated as having a prevalence under 1 in 2,000 by Orphanet) in the UK Biobank. We mapped ICD-10 codes to Orpha numbers through a consensus process to ensure they identify specific rare diseases, then used SnpEff to annotate rare (JAK2 V617F mutation, for example $p=5.30x10^{-40}$ in primary myelofibrosis. We also confirmed previously reported mutations for *HBB* in beta-thalassemia ($p=7.34x10^{-12}$) and *F11* in congenital factor XI deficiency ($p=3.41x10^{-11}$). Novel (not reported in ClinVar) associations were identified between a loss of function variant in *CALR* and essential thrombocythemia ($p=1.59x10^{-13}$), as well as an *SRSF2* missense variant, predicted to be deleterious by SIFT, CADD and PROVEAN among others, and chronic myelomonocytic leukemia ($p=1.19x10^{-13}$). Together these findings demonstrate the utility of applying enhanced rare disease mapping to population biobanks for rare disease research and, as more exome sequenced samples become available for the UK Biobank and other resources, it will be important to harness this information to improve the diagnosis of rare diseases.

PrgmNr 2541 - Generation and quality control of the biobank-scale methylation dataset in the VAâ□□s Million Veteran Program (MVP)

View session detail

Author Block: **F. DONG**[1,2], B. D. Barnes[3], W. Zhou[4,5], Y. Shi[1], A. Sendamarai[1,2], B. M. Despard[1,2], D. Okuhara[1,2], B. R. Gorman[1,2], U. Saxena[1], X. Gong[1], P. Devineni[1], Y. Chen[1], C. Pan[6,7], K. Markianos[1], P. S. Tsao[6,7], S. Pyarajan[1,8]; [1]Ctr. for Data and Computational Sci., VA Boston Hlth.care System, Boston, MA, [2]Booz Allen Hamilton, Inc., McLean, VA, [3]Illumina, Inc., San Diego, CA, [4]Ctr. for Computational and Genomic Med., The Children's Hosp. of Philadelphia, Philadelphia, PA, [5]Dept. of Pathology and Lab. Med., Univ. of Pennsylvania, Philadelphia, PA, [6]Dept. of Med., Stanford Univ. Sch. of Med., Palo Alto, CA, [7]VA Palo Alto Hlth.Care System, Palo Alto, CA, [8]Dept. of Med., Brigham and Women's Hosp. and Harvard Sch. of Med., Boston, MA

Disclosure Block: **F. Dong:** None.

The Department of Veterans Affairs (VA) Million Veteran Program (MVP) is one of the worldâ□□s largest databases of genomic and clinical data. In addition to genotype data, the MVP project is developing multi-omic resources. Here we introduce the new MVP methylation dataset. We used Illuminaâ□□s Infinium Methylation EPIC Array to quantify methylation at over 850,000 genomic sites for ~50,000 human blood samples. To routinely prepare and analyze data at this scale, we developed a Dockerized, high-throughput SeSAMe-based quality control and preprocessing pipeline incorporating common preprocessing steps (Type I probe channel inference, noob background correction, and dye bias correction) and sensitive pOOBAH (P-value with out-of-band array hybridization)-based probe-level quality control. We determined optimal pOOBAH and sample call rate thresholds using an epigenome-wide association study (EWAS; see below for details) and found that pOOBAH 94% yield a suitable balance for data quality and sample rejection rate. SeSAMe also includes age, sex, and novel genotype inference methods beyond the standard 59-marker fingerprinting which we use for sample-level QC. We present the utility of genotype inference for sample identity verification in a biobank-scale cohort. We demonstrate the quality and scientific potential of this data set by performing EWAS for a model phenotype, Body Mass Index (BMI). We replicate previously published results including 201 significant associations and coefficient signs among 216 previously reported CpGs from 3 large scale BMI studies. We also provide insights on the correlation of methylation levels and BMI as a function of disease status, e.g. diabetes diagnosis and/or renal disease. In addition, we analyzed age acceleration, the difference between methylation-derived predicted age (353 marker Horvath model) and chronological age, as a function of disease diagnosis. For example, in HIV-positive patients we find an average acceleration consistent with previously reported results of 3-4 years. Our novel findings illustrate the potential of this dataset to elucidate the role of epigenetic variation in human disease, particularly when combined with MVPâ□□s large genomic and clinical datasets.

PrgmNr 2542 - GHGA - The German Human Genome-Phenome Archive - A new federated national infrastructure for sharing human genomes in a European framework

View session detail

Author Block: **C. Mertes**, N. Smith, J. Gagneur, GHGA consortium; Dept. of Informatics, Technical Univ. of Munich, Garching, Germany
Disclosure Block: **C. Mertes:** None.
Genome sequencing and other omics technologies are among the most prominent and high-volume data sources in the life sciences, with major applications in basic biology, translational research, and medicine. Clinical omics profiling of patients is expected to dominate large-scale data generation in the near future, providing unprecedented opportunities for use of these data in research. While initiatives exist to harmonize phenotypic data, in particular medical health records, there is a lack of infrastructure for FAIR omics data within Germany. While these data are already generated in large scale across Germany, legal, ethical, and technical hurdles currently preclude managed access and data reuse for research at a national and international level.
Here we present the German Human Genome-Phenome Archive (GHGA) - an initiative to overcome these hurdles and balance the needs of data usage and data protection. Integrated into the German National Research Data Infrastructure (NFDI) and the federated European Genome-Phenome archive (EGA), GHGA will integrate existing and future omics data resources and link them to phenotypic information. With its 21 participating institutes, it will enable major scientific avenues through the delivery of harmonized molecular (meta-) data from large cohorts and create an invaluable bridge between biomedical research and healthcare, opening the door for scientists in and beyond Germany to participate in key international research networks.
GHGA activities will build on and extend existing, reliable, and secure high-performance computing infrastructures established by members of the consortium. A network of data hubs directly connected to the major data generators in Germany will handle the data in a federated manner. Using cloud technologies, this distributed infrastructure will be accessible to researchers in an integrated and seamless manner based on GA4GH standards. Researchers will have controlled access to raw sequence data as well as analysis results generated using harmonized, internationally recognized analysis workflows. The initial focus will be on seed communities that drive the national efforts for research and clinical sequencing at scale - rare diseases, oncology, and epidemiology. GHGA plans to meet its goals within a five year milestone based timeline until 2025. The consortium will drive open science solutions that are fully aligned with ELIXIRâ□□s and EGAâ□□s federation strategies. To ensure quality and comparability with international standards, GHGA will maintain FAIR data standards and engage with projects such as GA4GH to foster international data exchange in current and upcoming studies.

PrgmNr 2543 - Natural-language processing and machine learning can systematically infer tissue annotations of omics samples by modeling their unstructured metadata

View session detail

Author Block: **A. Krishnan**, N. T. Hawkins, M. J. Maldaver, A. Yannakopoulos, L. A. Guare; Michigan State Univ., East Lansing, MI
Disclosure Block: **A. Krishnan:** None.
Currently, there are >1.3 million human -omics samples from >26,000 datasets that are publicly available in repositories such as EBI ArrayExpress and NCBI GEO. These samples capture cellular responses of diverse human tissues/cell-types under thousands of conditions, making these published data invaluable for other researchers to reuse to answer new questions. However, currently, they cannot easily find the samples of interest because important sample attributes such as tissue of origin are routinely described using non-standard terminologies that are, in turn, buried within free-text descriptions.
To address this major problem, we have developed a new approach, NLP-ML, that combines natural language processing (NLP) and machine learning (ML) to annotate samples to their tissue-of-origin solely based on their unstructured text descriptions. NLP-ML works by creating numerical representations of sample descriptions and using these representations as features in a supervised ML classifier that predicts tissue/cell-type term annotations.
NLP-ML is the first method that takes a sophisticated approach to annotate omics samples using sample description that lacks any structure. Our approach significantly outperforms an advanced graph-based reasoning annotation method (MetaSRA) and a baseline exact string matching method (TAGGER). The trained NLP-ML tissue models capture biologically meaningful signals in text. Models of anatomically related tissues are similar to each other and can correctly classify tissue-associated biological processes and diseases based on their text descriptions alone. NLP-ML models are nearly as accurate as models based on gene-expression profiles in predicting sample tissue annotations but have the distinct capability to classify samples from any omics experiment type just based on their text metadata.
NLP-ML is implemented in the software Txt2Onto https://github.com/krishnanlab/txt2onto that contains a Python utility for classifying any unstructured text to terms in a tissue ontology, along with pretrained models, demo scripts, and extensive documentation. We are now using this method to annotate and distribute tissue/cell-type annotations of >1 million human omics samples.

PrgmNr 2544 - NDD-CNV Portal: Facilitating genetic test interpretation, research and education for Neurodevelopmental CNVs

View session detail

Author Block: **M. Macnee**[1], E. PÃ©rez-Palma[2], T. BrÃ¼nger[1], C. KlÃ¶ckner[3], A. Stefanski[4], P. May[5], J. Lemke[3], B. Shah[6], Y. DeWoody[7], V. Vogel-Farley[8], G. A. Logsdon[9], E. E. Eichler[9,10], D. Lal[4,11,1], Commission on Novel Technologies for Neurodevelopmental CNVs; [1]Cologne Ctr. for Genomics (CCG), Med. Faculty of the Univ. of Cologne, Univ. Hosp. of Cologne, Cologne, Germany, [2]Univ. del Desarrollo, Centro de GenÃ©tica y GenÃ³mica, Facultad de Med. ClÃnica Alemana., Santiago, Chile, [3]Inst. of Human Genetics, Univ. of Leipzig Med. Ctr., Leipzig, Germany, [4]Genomic Med. Inst. & Epilepsy Ctr., Cleveland Clinic, Cleveland, OH, [5]Bioinformatics Core, Luxembourg Ctr. for Systems Biomedicine, Univ. of Luxembourg, Esch-sur-Alzette, Luxembourg, [6]Project 8p Fndn., New York, NY, [7]Ring14 USA and Ring14 Intl., Midland, TX, [8]Dup15q Alliance, Highland Park, IL, [9]Dept. of Genome Sci., Univ. of Washington Sch. of Med., Seattle, WA, [10]Howard Hughes Med. Inst., Univ. of Washington, Seattle, WA, [11]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA
Disclosure Block: **M. Macnee:** None.

Copy number variants (CNVs) can cause a spectrum of neuropsychiatric and neurodevelopmental disorders (NDDs). However, variant interpretation, including phenotype driver and modifier gene identification, still represent a challenge, even for experts. Currently, clinical, genetic, and molecular data about CNVs are not connected and distributed across registries, databases, and the literature. To overcome current limitations, we are developing the NDD-CNV Portal, an interactive website that displays expert-curated CNV datasets alongside biomedical annotations, user-friendly analytics, and educational resources. In this ongoing project, we aggregated clinical, genomic, transcriptomic and proteomic data from 30 patients with pathogenic 8p (n=10), Ring14 (n=5), and Dup15 (n=15) CNVs from patient collaboratives. In addition, we collected and curated additional pathogenic and population CNVs from databases such as the UK-Biobank, gnomAD, ClinVar, and DECIPHER. To enable exploration of CNVs, we developed interfaces that visualize overlap of patient CNVs with population CNVs, genomic regulatory elements, disease-associated genes, SNVs, GWAS hits and >20 gene-level features such as (inter-and intraspecies) sequence constraint metric, dosage sensitivity, tissue, and cell-type level expression. By combining gene-level features, we rank genes in the order of most likely responsible genes and provide enrichment analyses of phenotype, functional, and pathway annotations. In addition to the search based on genomic position, the user can also perform disorder-specific and across-disorder genotype-phenotype analyses. We designed the NDD-CNV Portal for three user scenarios: i) Education of NDD-CNV-related diseases, ii) Expert-level variant interpretation using CNV guideline-based variant pathogenicity classification tools, and iii) Research using interactive tools and visualizations to explore the rich source of interconnected data and annotations. The NDD-CNV Portal infrastructure is scalable and will integrate novel data types. As such, it has the potential to transform variant interpretation, research, and education for neurodevelopmental CNV-associated disorders. We are actively looking for collaborators contributing to this project.

PrgmNr 2545 - SCN Portal for epilepsy genetics: Facilitating genetic test interpretation, research and education for sodium channel disorders

View session detail

Author Block: **T. BrÃ¼nger**[1], E. PÃ©rez-Palma[2], K. M. Johannesen[3,4], C. KlÃ¶ckner[5], M. Macnee[1], A. Stefanski[6], I. Helbig[7,8,9,10], S. Zuberi[11,12], P. May[13], J. Lemke[5], J. Q. Pan[14], A. L. George, Jr[15], R. S. MÃ¸ller[3,4], A. Brunklaus[11,12], D. Lal[16,17]; [1]Cologne Ctr. for Genomics, Univ. of cologne, Cologne, Germany, [2]Univ. del Desarrollo, Centro de GenÃ©tica y GenÃ³mica, Facultad de Med. ClÃnica Alemana, Santiago, Chile, [3]Danish Epilepsy Ctr., Filadelfia, Dianalund, Denmark, [4]Dept. for Regional Hlth.Res., Univ. of Southern Denmark, Odense, Denmark, [5]Inst. of Human Genetics, Univ. of Leipzig Med. Ctr., Leipzig, Germany, [6]Genomic Med. Inst., Lerner Res. Inst., Cleveland Clinic, Cleveland, OH, [7]Children's Hosp. of Philadelphia, Philadelphia, PA, [8]The Epilepsy NeuroGenetics Initiative (ENGIN), Children's Hosp. of Philadelphia, Philadelphia, PA, [9]Dept. of BioMed. and Hlth.Informatics (DBHi), Children's Hosp. of Philadelphia, Philadelphia, PA, [10]Dept. of Neurology, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [11]The Paediatric NeuroSci.s Res. Group, Royal Hosp. for Children, Glasgow, United Kingdom, [12]Inst. of Hlth.and Wellbeing, Univ. of Glasgow, Glasgow, United Kingdom, [13]Luxembourg Ctr. for Systems Biomedicine, Univ. of Luxembourg, Esch-Belval, Luxembourg, [14]Stanley Ctr. of Psychiatric Res., Broad Inst. of Harvard and MIT, 75 Ames Street, Cambridge, MA, [15]Dept. of Pharmacology, Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [16]Genomic Med. Inst. & Epilepsy Ctr., Cleveland Clinic, Cleveland, OH, [17]Stanley Ctr. of Psychiatric Res., Broad Inst. of Harvard and MIT, Cambridge, MA
Disclosure Block: **T. BrÃ¼nger:** None.

Pathogenic variants affecting the voltage-gated sodium channel (SCN) encoding genes *SCN1A*, *SCN2A*, *SCN3A*, and *SCN8A* cause a spectrum of neurodevelopmental disorders with or without epilepsy. Variant interpretation represents a challenge, even for SCN gene experts. Currently, clinical, genetic, and molecular data about SCN genes are not connected and distributed across registries, databases, and the literature. To overcome current limitations, we developed the SCN-Portal, which contains the largest expert-curated dataset of sodium channel disorders alongside user-friendly analytics and educational resources. Presently, we aggregated from 1849 patients with pathogenic variants in *SCN1A, SCN2A, SCN3A*, and *SCN8A* published and unpublished clinical and genetic data. We collected and curated electrophysiological data for 425 SCN gene variants from the literature and generated molecular read-outs for additional *SCN1A* and *SCN2A* variants. These data were connected with protein structures, and annotated with >40 protein features characterizing the location, physicochemical properties, and interactions of the variants within 3D protein structures. Several annotations, such as those for the size, location, and corresponding physiochemical properties of the pore, are novel and have been calculated using bioinformatic tools. Using our previously developed paralog framework, information from each SCN gene was transferred *in silico* to the other SCN genes. We designed the SCN-Portal for three user scenarios: i) Education: We provide a rich source of information in >10 languages for all known SCN1/2/3/8A-related diseases; ii) Access to expert-level variant interpretation: We provide novel expert-curated web applications that enable variant pathogenicity classification and disease trajectory prediction; iii) Research resource: Throughout the SCN-Portal, we provide novel research tools to explore our extensive and unique biomedical data resource. Users will be guided by tutorials with illustrative examples and will not require programming skills. The SCN-Portal infrastructure is scalable and, with increasing data, may transform variant interpretation, research and education for *SCN1A*, *SCN2A*, *SCN3A* and *SCN8A*. The SCN-Portal will be hosted at http://scn-portal.broadinstitute.org. We are actively looking for collaborators contributing to this project.

PrgmNr 2546 - Strategies to increase the availability of summary statistics in the GWAS Catalog

View session detail

Author Block: **A. Buniello**[1,2], J. Hayhurst[1,2], A. Abid[1], M. Cerezo[1], M. Ghoussaini[3,2], P. Hall[4], S. John[1], E. Lewis[1], J. A. MacArthur[1], A. MacMahon[1], A. Mosaku[1], S. Ramachandran[1], S. Saverimuttu[1], E. Sollis[1], F. Cunningham[1], P. Flicek[1], L. Hindorff[4], L. Harris[1], I. Dunham[1,2], P. Helen[1]; [1]European Molecular Biology Lab., European Bioinformatics Inst., Hinxton, Cambridge, United Kingdom, [2]Open Targets, Hinxton, Cambridge, United Kingdom, [3]Wellcome Trust Sanger Inst, Cambridge, United Kingdom, [4]Div. of Genomic Med., Natl. Human Genome Res. Inst., NIH, Bethesda, MD

Disclosure Block: **A. Buniello:** None.

The NHGRI-EBI GWAS Catalog is a central resource for the Genome Wide Association Study (GWAS) community, demonstrating the benefit of expert data curation and integration of full p-value GWAS summary statistics into a central repository for variant-trait associations. The Catalog aims to make GWAS data FAIR (Findable, Accessible, Interoperable and Re-usable), while serving as a starting point for investigations to identify causal variants, calculate disease risk, understand disease mechanisms and establish targets for novel therapies. The GWAS Catalog summary statistics repository is one of the largest, most visited and most frequently updated resources of its kind - now hosting full p-value results from at least 8,232 independent analyses (from 480 publications), which account for a total of 25,000 datasets from a wide variety of traits. Summary statistics can easily be accessed and downloaded from the GWAS Catalog FTP site or via a dedicated summary statistics API. In an aim to facilitate data sharing and interoperability, the GWAS team have recently released a web-based deposition interface to support scalable author submission of summary statistics and metadata from published and pre-published (e.g. submitted at the time of journal submission upon request from the reviewers) GWAS. The Nature Journals Group, for example, are proactively supporting an open access policy for their published GWAS data by adding the GWAS Catalog in their list of official summary statistics repositories. Currently, more than 80% of GWAS results are submitted to the Catalog pre-publication. The expansion of the GWAS Catalog summary statistics repository has required a considerable outreach effort from the Catalog data team and the main stakeholders, including funders and top journals. Concerns about data privacy and misuse, technical challenges and lack of standards are amongst the barriers to sharing the Catalog team encountered at the beginning of this journey. Following our successful 2020 workshop on GWAS summary statistics standards and sharing, we are continuing to work with the community to remove barriers to sharing. The rate of summary statistics data sharing, in fact, still noticeably differs among different genetics cohorts and therapy areas, and is particularly low in the cancer genetics community. Over the next few months, we will be focusing our community engagement efforts on the identification of strategies to maximise diversity of data shared in the Catalog, including underrepresented populations and traits. Furthermore, we will gather final recommendations on data and metadata standards to report and share GWAS results.

PrgmNr 2547 - The Brazilian Rare Genomes Project: the largest whole genome sequencing effort for rare disorders in South America

View session detail

Author Block: **J. OLIVEIRA FILHO**[1], T. Almeida[1], M. Cervato[1], M. Caraciolo[1], N. Zurro[1], G. Colichio[1], B. Mascaro[1], M. Nobrega[1], K. Chen[2], J. R. Ceroni[1], E. Perrone[1], L. Virmond[1], C. Quaio[1], C. A. Moreno[1], E. Alves[1], R. Siqueira[1], A. Coelho[1], R. Minillo[3]; [1]Hosp. Israelita Albert Einstein, Sao Paulo, Brazil, [2]SÃ£o Paulo, Brazil, [3]Hosp. Israelita Albert Einstein, SÃ£o Paulo, Brazil
Disclosure Block: **J. Oliveira filho:** None.
Rare diseases together comprise a group of 9,603 distinct entitites and have an estimated cumulative prevalence of 1.5 to 6.2% of the population. This represents, in Brazil, a total number of 3.2 to 13.2 million individuals, with a great impact on health services. Especially challenging is the definitive diagnosis of these patients, leading to true pilgrimages within the health system, with great delay and consequent increase in morbidity and mortality. The Rare Genomes Project is an initiative from the Albert Einstein Jewish Hospital in SÃ£o Paulo, in partnership with the National Health Ministry. It was initiated in 2020, with the aim of sequencing 7755 patients with rare disorders seen in the Brazilian public health system until the end of 2023. These patients are being recruited from participating centers spread across the country, from north to south. In 2020, the efforts were directed to building the sequencing, IT and bioinformatics infrastructure to support the project. So far, 1486 probands were recruited from 18 predefined disease groups, and whole genome sequencing was performed in over 1300 patients to date. Preliminary results are show in another abstract. Data collection is being done electronically, using Phenotips and Redcap. Sequencing protocols include PCR-free whole genome sequencing in Illumina Novaseq 6000 equipment , and bioinformatic processing is being performed using Dragen and in-house developed tools. Variant data will be deposited in Clinvar, and HPO terms with gene-level findings of negative cases will be shared via Matchmaker Exchange. The samples will be stored in a biobank, to accelerate future research in the field. This project will also allow the creation of the largest Brazilian genetic database of patients with RD and provide subsidies to assist in the implementation of genetic tools for future use in the public health system.

PrgmNr 2548 - Towards robust clinical genome interpretation: developing a consistent terminology to characterise disease-gene relationships for harmonised curation - an updated structural variant ontology for the NGS era

View session detail

Author Block: **A. M. Roberts**[1,2,3], M. DiStefano[4,5], E. R. Riggs[6], F. Cunningham[7], H. V. Firth[8,9], A. Benet-PagÃ¨s[10], C. L. Martin[11], C. Lowther[12,5,13], J. T. Den Dunnen[14], J. R. Vermeesch[15], R. M. Kuhn[16], M. J. Landrum[17], M. Good[4], T. Bingaman[4], L. Paul[18], M. Hughes[4], E. Wilcox[19], H. McCurry[20], J. S. Ware[1,21], H. L. Rehm[12]; [1]Natl. Heart & Lung Inst., Imperial Coll. London, London, United Kingdom, [2]Great Ormond Street Hosp., London, United Kingdom, [3]Royal Brompton & Harefield Hosp., London, United Kingdom, [4]Geisinger, Danville, PA, [5]Med. and Population Genetics Program, Broad Inst. of MIT and Harvard,, Cambridge, MA, [6]Geisinger Hlth.System, Huntsville, AL, [7]EMBl-EBI, Hinxton, Cambridgeshire, United Kingdom, [8]Wellcome Sanger Inst., Hinxton, United Kingdom, [9]Cambridge Univ. Hosp., Camberidge, United Kingdom, [10]UCSC Genomics Inst., Santa Cruz, CA, [11]Geisinger, Lewisburg, PA, [12]Massachusetts Gen. Hosp., Boston, MA, [13]Harvard Med. Sch., Cambridge, MA, [14]Rotterdam, Netherlands, [15]Univ. Hosp. of Leuven, Leuven, Belgium, [16]Univ. of California Santa Cruz, Santa Cruz, CA, [17]NCBI/NIH, Bethesda, MD, [18]Autism & Dev.al Med. Inst., Geisinger, Lewisburg, PA, [19]Broad Inst., Cambridge, MA, [20]Cambridge, MA, [21]Royal Brompton & Harefield Hosp.,, London, United Kingdom

Disclosure Block: **A.M. Roberts:** None.

Clinical application of genomic data is reliant on robust, curated associations between locus, genotype, mechanism, and disease phenotypes. Members of the Gene Curation Coalition (GenCC; www.thegencc.org) are working to harmonise approaches to ensure gene-level curated resources are comparable and interoperable, to streamline gene curation efforts and facilitate interpretation of variants. The GenCC includes members of Clinical Genome Resource (ClinGen), DECIPHER/Gene2Phenotype/Transforming Genetic Medicine Initiative (DECIPHER/G2P/TGMI), Medline Plus Genetics (formerly Genetics Home Reference GHR), Genomics England PanelApp (PanelApp), PanelApp Australia, Online Mendelian Inheritance in Man (OMIM), HUGO Gene Nomenclature Committee (HGNC), Orphanet, and private sector partners including clinical testing laboratories.

Little consensus exists on the use of public ontologies for annotation of structural variants (SV); most expert groups use an in-house system.

Here we present an approach to standardize terminology for SV annotation.

Aim: to harmonise description of SV from NGS data (ie WGS) with regard to a single gene, enabling more accurate data aggregation and assumptions about the mechanism through which these variants can cause disease.

Intended use:

1. Manual curation of disease-associated variant class and functional consequence in disease-gene pairs
2. Manual annotation and interpretation of variants identified in patients in the clinical and research setting
3. Consistent output of automated consequence annotation pipelines eg Ensemblâ□□s Variant Effect Predictor (VEP)
4. Matching SV consequences to curated gene-disease mechanisms for variant filtering in genome annotation pipelines

Our SV working group includes individuals with experience in identification and evaluation of structural variants in both the clinical and research settings, bioinformatics and curation tool development, as well as individuals representing current nomenclature standards.We drafted a hierarchical ontology to describe the content, orientation and location of potentially clinically relevant SVs with regard to their impact on a single gene.

The ontology was piloted by manual annotation of SVs by a team of biocurators on 59 variants from 25 ClinGen dosage curations, and feedback incorporated. These clinical variants are enriched for transcript deletions. Further use testing is ongoing using SV from the GnomAD SV dataset including translocations and transcript duplications.
Working with Sequence Ontology we will integrate and update these terms such that these will be available centrally and with an associated public definition.

PrgmNr 2549 - *De novo* pathogenic variants predominate in exomes of a cohort of patients with intellectual disabilities and/or congenital anomalies. Preliminary results from a pilot program for Rare Undiagnosed Diseases in Chile

View session detail

Author Block: **G. M. Repetto**[1], L. M. Martin[1], C. Lagos[1], G. Encina[2], B. Rebolledo-Jaramillo[2], C. Poli[1], G. Lay-Son[3], E. San Martin[4], T. MuÃ±oz[1], R. Armisen[1], S. Fischer[1], F. Brito[1], S. Sifaqui[1], L. Noguera[1], M. Rojas[1], E. Gutierrez[1], J. E. Posey[5], J. R. Lupski[6]; [1]Clin Alemana- Univ Desarrollo, Santiago, Chile, [2]Clin Alemana-Univ del Desarrollo, Santiago, Chile, [3]P Univ Catolica de Chile, Santiago, Chile, [4]Hosp. Guillermo Grant Benavente, Concepcion, Chile, [5]Baylor Coll. of Med., Houston, TX, [6]Baylor Col Med, Houston, TX

Disclosure Block: **G.M. Repetto:** None.

Rare Diseases (RD) affect populations worldwide. Countries with limited genomic resources face challenges to implement RD programs, such as availability of clinical genetics and genomics specialists, sequencing and bioinformatics resources, and lack of insurance coverage for genomic testing. To build local capacities to address these gaps, we developed a pilot program for clinical evaluation and exome sequencing (ES) of individuals with congenital anomalies and/or intellectual disabilities and unknown diagnoses in Chile. We used different NGS alternatives, from in-house singleton clinical exome sequencing (cES) (approximately 5000 known disease-causing genes) to outsourced research ES strategies, proband-only and trios, and bioinformatics analysis using in-house pipelines and commercially available tools. These processes were followed by multidisciplinary team variant interpretation and Sanger confirmation in the proband and parents if available. To date, 38 patients have participated in the study. cES was used in the first 15 participants, 3 of them (20%) had a heterozygous *de novo*, likely causative variant in *NFIX*, *TCF4* and *AHDC1*. ES-proband only was completed in 18 participants (11 with prior non-informative cES) and identified a causative variant in 8 (44.4%), including 5 *de novo* or likely *de novo* SNVs (*DDX3, ARID2, KAT6A, PUF60, SOX10*), 2 recessive conditions (*TECPR2, UBE3B*) and 1 *de novo* CNV (6q24q25 deletion). ES-trio was performed in another 16 families, resulting in 8 diagnoses (50%): 4 *de novo* SNVs (*MAF, TBC1D2B, SPTBN2, SMARCA4*), 2 inherited dominant variants (*NOG* and *EFTUD2*), 1 autosomal recessive condition (*TK2*), and 1 CNV (Xq13q21 duplication). Overall diagnostic yield was 50%, each participant had a different molecular finding and *de novo* variants accounted for 14/19 cases with a diagnosis (82%), providing information for genetic counseling. These results show the feasibility of developing RD programs in limited-resource settings, using combinations of local and outsourced capacities. Future challenges include scaling up, performing cost-utility analysis, understanding patient perceptions, assessing changes in clinical management after molecular diagnosis to guide clinical implementation in the Chilean health care system. Research funded in part by ANID-Chile grants FONDECYT 1171014 and 1211411, FONDEQUIP EQM150093, Redes 180047, US National Human Genome Research Institute (NIH) Baylor Hopkins Centers for Mendelian Genomics, and a donation from Child Health Foundation, Birmingham, AL

PrgmNr 2550 - *De novo* variants in *TCF4* with a suspected gain-of-function mechanism are responsible for a new malformative disease without intellectual disability

View session detail

Author Block: **E. Colin**[1,2], M. De Sarlo[3], J. Paccaud[1], C. Gabellini[4], D. GeneviÃ¨ve[5], D. Wegner[6], M. Shinawi[7], J. Delanne[1,8], F. Tran Mau-Them[1,9], M. Alders[10], L. Menke[11], S. Nambot[1,8], A-S. DenommÃ©-Pichon[1,9], A-L. Bruel[1,9], J. Thevenon[1,8,9], J-B. RiviÃ¨re[1,9], A. Foroutan[12], H. McConkey[13], J. Kerkhof[14], M. Chevarin[1,9], Y. Duffourd[1,9], C. Philippe[1,9], K. Leblanc[15], C. Thauvin-Robinet[1,9], P. Blackburn[16], The Solve-RD consortium, B. Sadikovic[17], L. Faivre[1,8], M. Ori[18], A. Vitobello[1,9]; [1]UMR-Inserm 1231 GAD team, GÃ©nÃ©tique des Anomalies du dÃ©veloppement, UniversitÃ© de Bourgogne Franche-ComtÃ©, FHU-TRANSLAD, Dijon, France, [2]Service de GÃ©nÃ©tique MÃ©dicale, CHU d'Angers, Angers, France, [3]Dept. of Biology, Univ. of Pisa, Pisa, Italy, [4]Unit of Cell and Dev.al Biology, Dept. of Biology, Univ. of Pisa, Pisa, Italy, [5]DÃ©partement de gÃ©nÃ©tique mÃ©dicale, Maladies rares et mÃ©decine personnalisÃ©e, CHU de Montpellier, Montpellier, France, [6]Dept. of Pediatrics, Div. of Genetics and Genomic Med., Washington Univ. Sch. of Med., St. Louis, MO, [7]Washington Univ. Sch. of Med., St Louis, MO, [8]Ctr. de GÃ©nÃ©tique et Ctr. de rÃ©fÃ©rence Â« Anomalies du DÃ©veloppement et Syndromes Malformatifs Â», HÃ´pital dâ□□Enfants, Ctr. Hosp.ier Univ.ire de Dijon, Dijon, France, [9]UF Innovation en diagnostic gÃ©nomique des maladies rares, Ctr. Hosp.ier Univ.ire de Dijon, FHU-TRANSLAD, Dijon, France, [10]AMC, Amsterdam, NH, Netherlands, [11]Dept. of Pediatrics, Emma Children's Hosp., Amsterdam UMC, Univ. of Amsterdam, Amsterdam, Netherlands, [12]Molecular Diagnostics Div., Pathology and Lab. Med., Univ. of Western Ontario, London, ON, Canada, [13]Molecular Diagnostics Div., Pathology and Lab. Med., Univ. of Western, London, ON, Canada, [14]London Hlth.Sci. Ctr., London, ON, Canada, [15]Undiagnosed Diseases Network, Harvard Med. Sch., Boston, Boston, MA, [16]St. Jude Childrenâ□□s Res. Hosp., Memphis, TN, [17]LHSC, London, ON, Canada, [18]Dipartimento di Biologia, UniversitÃ di Pisa, Pisa, Italy

Disclosure Block: **E. Colin:** None.

TCF4 loss-of-function variants cause Pitt-Hopkins syndrome (PTHS), associating intellectual disability (ID), wide mouth, distinctive facial features, and intermittent hyperventilation followed by apnea. Pathogenic missense variants are primarily clustered within the C-terminal HLH domain required for dimerization and DNA binding. Variants located elsewhere may be associated with mild non-syndromic ID. Using exome sequencing, we identified *de novo* missense variants in *TCF4* in three individuals who did not show typical PTHS hallmarks. The variants affect ultra-conserved amino acids within or close to the C-terminal HLH domain. These individuals showed consistent phenotypic features associating cranio-facial dysmorphism, limb anomalies and growth failure without ID. The main facial characteristics include an abnormal skull shape, wide forehead, sparse eyebrows, epicanthus, anteverted nares, short columella, micrognathia, low-set ears with external ear malformation. All individuals have lacrymal duct obstruction. Limbs malformations include camptodactyly of the fingers and toes, clinodactyly of the 5th finger, syndactyly and nail hypo/dysplasia. In two individuals, growth failure required growth hormone therapy. Neurological examination and psychometric assessment were normal. In order to study in vivo the role of the de novo *TCF4* missense variants, we are exploiting two vertebrate model systems: Xenopus laevis and Danio rerio (zebrafish). We firstly confirmed the conservation of the gene expression profile of tcf4 in Xenopus and zebrafish craniofacial development. Preliminary results in *Xenopus* embryos of the wild-type and the mutated forms of *TCF4* indicate an effect on cartilage development. Furthermore, we are also performing *in silico* modeling prediction as well as genome-wide DNA methylation, chromatin immuno-precipitation and mRNA sequencing in patient-derived samples. Overall, we report a new clinical entity associated with suspected gain-of-function variants in *TCF4*, distinct from PTHS, with

facial and limb dysmorphism, growth failure and no ID. The description of other individuals will help to further delineate the phenotypic description. The generation of *in silico, in vitro* data and *in vivo* models for *TCF4* variants will allow us to define their role during development and to generate new platform for drug screening approaches.

PrgmNr 2551 - Biallelic loss-of-function variants in *CACHD1*, encoding an Î±2Î´-like voltage-gated calcium channel regulator, cause a neurodevelopmental, craniofacial, and genitourinary syndrome

View session detail

Author Block: **M. Scala**[1,2], C. Beneteau[3,4], K. Khan[5], B. CognÃ©[3,6], S. von Hardenberg[7], A. Khan[5], M. Joubert[4,8], L. Fievet[9], M. Musquer[4,8], C. Le Vaillant[10], J. Korda Holsclaw[9], D. Lim[11,12], A. Accogli[1], M. Mancardi[1,13], A. Torella[14,15], L. Nobili[1,13], P. Striano[1,2], V. Nigro[15,14], V. Capra[16], S. BÃ©zieau[3,6], E. Davis[5,17], M. Wells[18,19]; [1]Univ. of Genoa, Dept. of NeuroSci.s, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Hlth., Genoa, Italy, [2]Pediatric Neurology and Muscular Diseases Unit, IRCCS Istituto Giannina Gaslini, Genoa, Italy, [3]CHU Nantes, Dept. of Med. Genetics, Nantes, France, [4]CHU Nantes, UF of FÅtopathology and genetics, Nantes, France, [5]Advanced Ctr. for Translational and Genetic Med., Stanley Manne Children's Res. Inst., Ann & Robert H. Lurie Children's Hosp. of Chicago, Chicago, IL, [6]INSERM, CNRS, UNIV Nantes, l'institut du thorax, Nantes, France, [7]Dept. of Human Genetics, Hannover Med. Sch., Hannover, Germany, [8]CHU Nantes, Dept. of Anatomical Pathology, Nantes, France, [9]Ctr. for Human Disease Modeling, Duke Univ. Med. Ctr., Durham, NC, [10]CHU Nantes, Dept. of Obstetrics and Gynecology, Nantes, France, [11]Dept. of Clinical Genetics, Birmingham Women's and Children's NHS Fndn. Trust & Birmingham Hlth.Partners, Birmingham, United Kingdom, [12]Dept. of Med., Univ. of Birmingham, Birmingham, United Kingdom, [13]Child Neuropsychiatry Unit, IRCCS Giannina Gaslini Inst., Genoa, Italy, [14]Dept. of Precision Med., Univ. of Campania 'Luigi Vanvitelli', Naples, Italy, [15]Telethon Inst. of Genetics and Med., Pozzuoli, Italy, [16]Med. Genetics Unit, IRCCS Giannina Gaslini Inst., Genoa, Italy, [17]Dept. of Pediatrics, Feinberg Sch. of Med., Northwestern Univ., Chicago, IL, [18]Harvard Stem Cell Inst., Dept. of Stem Cell and Regenerative Biology, Harvard Univ., Cambridge, MA, [19]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA

Disclosure Block: **M. Scala:** None.

Background Voltage-gated calcium (CaV) channels are essential components and key functional regulators of excitable cells. Neuronal N-type (CaV2.2) and P/Q-type (CaV2.1) channels are critical for presynaptic neuro-transmitter release. CaV Î±1 subunits form the channel pore, whereas the Î² and Î±2Î´ subunits are crucial regulators of trafficking and bio-physical properties of channel complexes. The putative cache (Ca^{2+} channel and chemotaxis receptor) domain containing 1 (CACHD1) protein structurally mimics Î±2Î´ proteins and modulates CaV2 and CaV3 activity and expression.

Methods Using gene matching platforms, we assembled a cohort of six affected individuals from four unrelated families who present with a hitherto unreported neurodevelopmental syndrome. Exome sequencing (ES) was performed to identify the underlying genetic cause. We established stable *cachd1* zebrafish mutants on transgenic reporter lines relevant to clinical features in humans and performed quantitative phenotyping. Histology was performed on E18.5 *Cachd1* knockout mouse embryos to assess neurodevelopmental and gross morphological defects.

Results The affected individuals from families 1-3 display global psychomotor delay, facial dysmorphism, oculo-auricular malformations, congenital genitourinary abnormalities, and anorectal malformations. Family 4 consists of two aborted male fetuses both showing urethro-renal obstruction with hydronephrosis, periauricular skin tags, dysmorphic facial features, and esophageal atresia. ES led to the identification of biallelic loss-of-function variants in *CACHD1* (NM_020925.4) segregating with disease in all pedigrees: c.1783-1G>A and c.2387+1G>A in family 1; c.261+2T>C and c.648delC; p.(Ile217SerfsTer13) in family 2; homozygous c.274dup; p.(Ile92Asnfs*52) in family 3; and c.277C>T; p.(Arg93*) and c.460C>T; p.(Arg154*) in family 4. Homozygous *cachd1* mutant zebrafish larvae recapitulated hallmark features of affected humans, including aberrant craniofacial patterning and significant dilatation of the distal pronephric tubule and cloaca in comparison to wild type.

Conclusion These findings support biallelic *CACHD1* loss of function variants as the cause of a novel

neurodevelopmental syndrome and expand the clinical spectrum of voltage gated calcium channel effectors.

PrgmNr 2552 - NSD1 gene variations associated to Sotos syndrome in humans and diverse selection within exons

View session detail

Author Block: **S. Aguiar**[1], V. Romero[1], S. Saenz[1], K. Hosomichi[2]; [1]Univ. San Francisco de Quito, Quito, Ecuador, [2]Univ. de Kanazawa, Kanazawa, Japan
Disclosure Block: **S. Aguiar:** None.
Introduction Sotos syndrome is an overgrowth disease that affects 1/40 000 newborns worldwide and is associated with variants in *NSD1.* Here we report a novel duplication in an Ecuadorian toddler, perform a systematic literature review of the condition and localize a hot spot for variations in humans and primates.
Methods To find the causal molecular variation, we ordered a panel for overgrowth syndromes and identified a novel duplication in the NSD1 gene. We performed a systematic literature review of all cases reported and identified an exon susceptible to nucleotide variations. Therefore, we localize the NSD1 exons in primates and compare the percentage of similarity and synonymous/non-synonymous variations among species.
Results and Discussion There are 304 variants reported cases with this genetic condition. Most of the variants are associated with the well-known clinical manifestation and physical appearance of the syndrome. Deletions result in a classical Sotos syndrome, however, duplication on the gene results in an opposite phenotype (microcephaly and short stature). The genotype in our patient reported a duplication, however the phenotype resembles a classic mild Sotos Syndrome.
Conclusion In summary, our patientâ□□s variant leads to a misfolded protein that might simulate the common *NSD1* nonsense mutation which affects the length of the protein. Most of reported variants occur in exon 5 which suggests a relaxed selection with a pathological phenotype in humans.

PrgmNr 2553 - Clinical delineation, sex differences and genotype-phenotype correlation in pathogenic *KDM6A* variants causing X-linked Kabuki syndrome Type 2

View session detail

Author Block: **V. Faundes**[1,2], S. Goh[2], R. Akilapa[3], H. Bezuidenhout[4], H. T. Bjornsson[5], L. Bradley[6], A. Brady[3], E. Brischoux-Boucher[7], H. G. Brunner[8], S. BULK[9], N. Canham[10], D. Cody[6], M. L. Dentici[11], M. C. Digilio[11], F. Elmslie[12], A. Fry[13], h. k. gill[14], J. Hurst[15], D. Johnson[16], S. Julia[17], K. Lachlan[18], R. R. Lebel[19], M. Byler[20], E. M. Gershon[21], E. G. Lemire[22], M. Gnazzo[23], F. R. Lepri[23], A. Marchese[24], M. McEntagart[25], J. M. McGaughran[26], S. Mizuno[27], N. Okamoto[28], C. Rieubland[29], J. Rodgers[26], E. Sasaki[6], E. Scalais[30], I. Scurr[31], M. Suri[32], I. van der Burgt[8], N. Matsumoto[33], N. Miyake[34], V. Benoit[35], D. J. Lederer[36], S. Banka[37]; [1]INTA, Univ. de Chile, Santiago, Chile, [2]Univ. of Manchester, Manchester, United Kingdom, [3]NW Thames Regional Genetics Service, Northwick Park Hosp., Harrow, United Kingdom, [4]Stellenbosch Univ., Cape Town, South Africa, [5]Johns Hopkins Univ, Baltimore, MD, [6]Dept. of Clinical Genetics, Childrenâ□□s Hlth.Ireland at Crumlin, Dublin, Ireland, [7]CHU Saint Jacques, BesanÃ§on, France, [8]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [9]LIEGE, Belgium, [10]Liverpool Ctr. for Genomic Med., Liverpool Womenâ□□s Hosp., Liverpool, United Kingdom, [11]Bambino GesÃ¹ Childrenâ□□s Hosp., Rome, Italy, [12]SW Thames Regional Genetics Service, St George's, Univ. of London, London, United Kingdom, [13]Inst. of Med. Genetics, Univ. Hosp. of Wales, Cardiff, United Kingdom, [14]BC Women's Hosp. and Hlth.Ctr., Vancouver, BC, Canada, [15]NE Thames Genetics Service, Great Ormond Street Hosp., London, United Kingdom, [16]Sheffield Clinical Genetics Service, Sheffield Children's NHS Fndn. Trust, Northern Gen. Hosp., Sheffield, United Kingdom, [17]Inserm U 1027 and Univ. Hosp. of Toulouse, Toulouse, France, [18]Wessex Clinical Genetics Service and Div. of Human Genetics, Princess Anne Hosp., Southampton, United Kingdom, [19]Fayetteville, NY, [20]Dept. of Pediatrics, Section of Med. Genetics, SUNY Upstate Med. Univ., Syracuse, NY, [21]Upstate Med. Univ., Syracuse, NY, [22]Riverside Estates, SK, Canada, [23]Bambino GesÃ¹ Children's Hosp., Rome, Italy, [24]Service de PÃ©diatrie, Ctr. Hosp.ier RÃ©gional de Namur, Namur, Belgium, [25]St. George`s Univ. Hosp NHS Fndn. Trust, London, United Kingdom, [26]Royal Brisbane & Women's Hosp, Brisbane, Australia, [27]Central Hosp., Aichi Dev.al Disability Ctr., Kasugai, Aichi, Japan, [28]Osaka Women's and Children's Hosp., Osaka, Japan, [29]Div. of Human Genetics, Dept. of Pediatrics, Inselspital, Bern Univ. Hosp., Univ. of Bern, Bern, Switzerland, [30]Dept. of Pediatric Neurology, Natl. Hosp., Luxembourg City, Luxembourg, [31]Clinical Genetics, Univ. Hosp. Bristol, Bristol, United Kingdom, [32]Nottingham Clinical Genetics Service, City Hosp. Campus, Nottingham, United Kingdom, [33]Yokohama City Univ Grad Sch Med, Yokohama, Kanagawa, Japan, [34]Yokohama City Univ, Yokohama, Japan, [35]Inst Pathology & Genetic, charleroi, Belgium, [36]Inst Pathologie et Genetique ASBL, Gosselies, Hainaut, Belgium, [37]Univ. of Manchester, Sale, United Kingdom

Disclosure Block: **V. Faundes:** None.

Introduction: Kabuki syndrome is one of the most recognisable multi-system congenital disorder. Type-2 Kabuki syndrome (KS2) is a highly variable X-linked condition caused by *KDM6A* variants, and is thought to be responsible for ~5% of KS cases. Its mutation spectrum and the genotype-phenotype correlations are poorly understood. Methods: Genetic and clinical details of new and published individuals with pathogenic *KDM6A* variants were compiled and analysed. Results: 61 distinct pathogenic *KDM6A* variants (50 truncating, 11 missense) from 80 patients (34 males, 46 females) were identified. Missense variants clustered in the TRP #2, #3, #7 and Jmj-C domains. Truncating variants were significantly more likely to be *de novo*. Thirteen individuals had maternally inherited variants and one had a paternally inherited variant. Neonatal feeding difficulties, hypoglycaemia, post-natal growth retardation, poor weight gain, motor delay, intellectual disability (ID), microcephaly, congenital heart anomalies, palate defects, renal malformations, strabismus, hearing loss, recurrent infections, hyperinsulinism, seizures, joint hypermobility and gastroesophageal reflux were frequent

clinical findings. Facial features of >1/3rd patients were atypical for KS. Males were significantly more likely to be born prematurely, have shorter stature and severe developmental delay/ID. We also highlight the overlaps and differences between the phenotypes of KS2 and KS1. Conclusion: This largest-ever KS2 series expands the *KDM6A* mutation spectrum, delineates the KS2 phenotype and demonstrates its sex and variant-dependent variability. These results will improve diagnosis for patients with KS2, especially those with inherited missense variants. In future these results will inform the development of evidence-driven management guidelines in KS.

PrgmNr 2554 - Diverse phenotype genotype characterization and novel *SOX9* variation in Campomelic Dysplasia

View session detail

Author Block: **E. Vasquez**[1], C. Calvache[1], V. Romero[1], K. Hosomichi[2]; [1]USFQ, Quito, Ecuador, [2]Univ. de Kanazawa, Kanazawa, Japan
Disclosure Block: **E. Vasquez:** None.
BackgroundCampomelic dysplasia is a rare disorder (1/200,000 births) that involves the development of the skeletal and genital systems with a 10% survival rate. Our patient is a 4-year-old (female assigned at birth) with campomelic dysplasia, female sex reversal, type 1 Arnold Chiari malformation and bilateral conductive hearing loss.
Methods and Results To find the causal molecular variation, we performed in trio whole exome sequencing analysis and identified a novel variation in the *SOX9* gene. The *SOX9* gene plays a key role in the development of the endochondral skeleton and male sexual differentiation. The variation we found is a novel non-sense variation in exon 1 caused by a change of guanine (G) to adenine (A) at position 344bp, replacing the normal synthesis of tryptophan to a premature stop codon at amino acid position 115.
There is a diversity of phenotypes and genotypes expressions related to sex reversal, life span and prognosis associated with the *SOX9* gene variations. We performed a literary review of cases reported and identified hot spot locations in the gene associated to the different clinical presentations. The position of the variation correlates with the diversity of clinical manifestations and the similarity between patients with variants in the same position. In our patient it translates to a decreased level of pulmonary and airway involvement, being part of the small percentage that manages to survive beyond the first days of birth.
ConclusionsWe identified the casual novel variation responsible for the phenotype of our patient and recognized locations withing the *SOX9* gene associated to sex reversal, life span and prognosis. Our study suggests that the relation between phenotypes and genotypes is essential for the adequate counseling for the patient and the family.

PrgmNr 2555 - Exome sequencing identifies a missense mutation in *MMP14* in autosomal dominant hypotrichosis

View session detail

Author Block: **A. Onoufriadis**[1], J. Shi[2], J. K. Simpson[1], C. Tziotzios[1], C. M. Stefanato[3], P. Lovell[4], C. Tierney[1], W. Scott[1], S. M. Lwin[1], M. A. Simpson[5], C. A. Jahoda[6], P. G. Robey[2], J. A. McGrath[1]; [1]King's Coll. London, London, United Kingdom, [2]NIH, Bethesda, MD, [3]Guyâ□□s and St Thomas' NHS Fndn. Trust, London, United Kingdom, [4]Viapath, Guy's Hosp., London, United Kingdom, [5]King S Coll. London, London, United Kingdom, [6]Durham Univ., Durham, United Kingdom

Disclosure Block: **A. Onoufriadis:** None.

Hypotrichosis simplex (HS) refers to a group of hereditary isolated alopecias which are characterized by diffuse hair loss that usually begins in early childhood. Inherited forms of hypotrichosis are genetically heterogeneous with 11 genes implicated so far, including *APCDD1, CDSN, EPS8L3, KRT74, RPL21, SNRPE* and *U2HR*, which are responsible for dominant forms of the disease. To identify the underlying genetic defect in an HS-affected multigenerational pedigree, we undertook whole-exome sequencing (WES) in 5 affected individuals and focused on shared heterozygous predicted protein altering substitutions and indels with a minor allele frequency (MAF) of less than 0.5% in public and in-house exome databases. This analysis highlighted two novel variants, of which only a missense variant (c.1655T>A; p.Val552Gly) in *MMP14* co-segregated with the disease. This variant maps to the transmembrane domain of the protein, possibly affecting membrane binding of MMP14 which in turn can affect the activity of other MMPs around hair follicles. To assess the functional impact of the c.1655T>A mutation on *MMP14* expression, we performed quantitative reverse transcriptase in real time PCR on RNA derived from whole scalp skin biopsy from an affected individual and a healthy control, which showed a significant reduction of *MMP14* expression as well as increased *MMP2* expression. To investigate further MMP14 pathology in the context of hypotrichosis, we obtained skin punch biopsies from Mmp14-deficient and wild-type mice and performed transmission electron microscopy which showed increased vesicular activity and dense collagen in the dermis in Mmp14-deficient skin consistent with disruption of extracellular matrix anchoring hair follicles. Of note, patchy hair loss has been observed in Mmp14-deficient mice. In summary, we present a multigenerational pedigree in which defective MMP14 leads to hypotrichosis and therefore expand the genetic causes of hair disorders.

PrgmNr 2556 - First identification of an individual with atypical Smith-Magenis syndrome without intellectual disability and non-coding structural variant

View session detail

Author Block: **A. MASSON**[1], M. TESSARECH[2,3], A. GUICHET[2], C. BRIS[2,3], Y. DUFFOURD[1,4], V. RAVEROT[5], J. PACCAUD[1], A. TRIMOUILLE[6], B. NAVET[2], S. KHIATI[3], A. ZIEGLER[2,3], V. PROCACCIO[2,3], L. FAIVRE[1,7], D. BONNEAU[2,3], C. THAUVIN-ROBINET[1,4], A. VITOBELLO[1,4], E. COLIN[2,3]; [1]UMR Inserm 1231, GAD team, GÃ©nÃ©tique des Anomalies du DÃ©veloppement, UniversitÃ© de Bourgogne Franche-ComtÃ©, FHU-TRANSLAD, Dijon, France, [2]Service de GÃ©nÃ©tique MÃ©dicale, Ctr. Hosp.ier Univ.ire d'Angers, Angers, France, [3]UniversitÃ© d'Angers, Ctr. Hosp.ier Univ.ire dâ□□Angers, Inserm, CNRS, MITOVASC, SFR ICAT, Angers, France, [4]UF Innovation en diagnostic gÃ©nomique des maladies rares, Ctr. Hosp.ier Univ.ire de Dijon, FHU-TRANSLAD, Dijon, France, [5]Ctr. de Biologie et de Pathologie Est, Hospices Civils de Lyon, Groupement Hosp.ier Est, LBMMS, Bron, France, [6]Service de GÃ©nÃ©tique MÃ©dicale, Ctr. Hosp.ier Univ.ire de Bordeaux, Bordeaux, France, [7]DÃ©partement de gÃ©nÃ©tique mÃ©dicale, Maladies rares et mÃ©decine personnalisÃ©e, Ctr. Hosp.ier Univ.ire de Montpellier, Montpellier, France

Disclosure Block: **A. Masson:** None.

RAI1 haploinsufficiency is involved in Smith-Magenis syndrome (SMS - OMIM 182290), associating fully penetrant traits such as mild to severe intellectual disability (ID), distinctive physical features, behavioral abnormalities, and sleep disturbance. We identified an individual presenting with an early obesity, a broad square-shaped face, midface retrusion, and an everted upper lip with a ''tented'' appearance, and sleep disturbance but without ID (VCI 116, FRI 106, WMI 94, PSI 111, Wechsler Intelligence Scale for Children-V Edition), whose clinical presentation was partially overlapping with SMS. This suspected clinical diagnosis led to the re-evaluation of exome sequencing and SNP array data, with the identification of an unreported deletion of 229 kb including the first two non-coding exons of RAI1 as well as part of its regulatory regions. Genome sequencing was performed to exclude any additional chromosomal rearrangements or aberrant events potentially accounting for the clinical presentation. The deletion extended into the adjacent intergenic region including part of the contiguous gene PEMT and containing a topologically associating domain (TAD) boundary, suggesting a possible fusion of the two contiguous TADs. We thus speculated that the atypical SMS could be due to a reshaped long-range interaction landscape of the locus, causing tissue-specific haploinsufficiency. To test this hypothesis we are currently performing mRNA sequencing coupled with whole genome chromosome conformation capture in primary fibroblasts and an in vitro
neural cell line to explore the epigenetic and transcriptional landscape associated with this variant. The identification of further patients carrying overlapping clinical features and molecular results will help us understanding the mechanisms implicated in the atypical presentation of our patient.

PrgmNr 2557 - Five new patients with pathogenic variants in*ZC4H2*expand the genotypic and phenotypic spectrum of*ZC4H2*-related disorder

View session detail

Author Block: **P. Wongkittichote**[1], T-I. Choi[2], K. Riley[3], D. D. Koeberl[3], C. E. Schwartz[4], V. Narayanan[5], K. Ramsey[5], K. Cheol-Hee[2], M. Shinawi[6]; [1]Dept. of Pediatrics, Div. of Genetics and Genomic Med., Washington Univ. Sch. of Med., St Louis, MO, [2]Chungnam Natl. Univ., Daejeon, Korea, Republic of, [3]Duke Univ. Sch. of Med., Durham, NC, [4]Greenwood, SC, [5]Translational Genomics Res. Inst., Phoenix, AZ, [6]Washington Univ. Sch. of Med., St Louis, MO
Disclosure Block: **P. Wongkittichote:** None.
ZC4H2 (MIM# 300897) is a nuclear factor involved in various cellular processes including proliferation and differentiation of neural stem cells, ventral spinal patterning and osteogenic and myogenic processes. Pathogenic variants in *ZC4H2* have been associated with Wieacker-Wolff syndrome (MIM# 314580), an X-linked neurodevelopmental disorder characterized by infantile spasms, arthrogryposis, development delay, hypotonia, feeding difficulties, poor growth, skeletal abnormalities and dysmorphic features. Zebrafish *zc4h2* null mutant fish were viable beyond embryogenesis, recapitulated the human phenotype, showed complete loss of *vsx2* expression in brain, and exhibited abnormal swimming and balance problems. Here we report 5 new (4 males and 1 female) patients with *ZC4H2*-related disorder from 4 unrelated families. Three of the four *ZC4H2* variants are novel: 2 are missense, designated as c.142T>A (p.Tyr48Asn) and c.558G>A (p.Met186Ile), 1 is nonsense variant, c.618C>A (p.C206*). One variant is a splice-site (c.225+5G>A). Bioinformatic analysis of these variants supported their pathogenicity. Four patients were on the severe spectrum of clinical findings associated with *ZC4H2*-related disorder, one of whom died at 7 years of age. The male patient harboring hemizygous p.Met186Ile has a relatively mild phenotype. Of note, 3/5 patients had a tethered cord requiring a surgical repair. To study the effect of the missense variants, we performed microinjection of human *ZC4H2*wild-type or variant mRNAs into one-cell stage *zc4h2* null mutant zebrafish embryos. The p.Met186Ile mRNA variant was able to partially rescue *vsx2* expression while p.Tyr48Asn mRNA variant was not. However, swimming and balance problems could not be rescued by either *ZC4H2* variant. These results suggest that the p.Met186Ile is a hypomorphic allele, which can explain the associated milder phenotype. Our work expands the genotypes and phenotypes associated with *ZC4H2*-related disorder and demonstrates that the zebrafish system is a reliable method to determine the pathogenicity of ZC4H2 variants. Larger cohorts of patients are needed to study genotype-phenotype correlations, the effect of additional hypomorphic variants, and the prevalence of tethered cord.

PrgmNr 2558 - Genotype-phenotype correlations in *PIK3CA*-related overgrowth spectrum (PROS) and overlapping phenotypes: a systematic review of 1007 patients with *PIK3CA* pathogenic variants

View session detail

Author Block: **D. Carli**[1], C. Leoni[2], M. Iacoviello[3], A. Mussa[1], C. Ranieri[3], A. Pantaleo[3], P. S. Buonuomo[4], B. Rosanna[3], G. B. Ferrero[5], D. Melis[6], S. Maitz[7], D. C. Loconte[3], A. Turchiano[3], M. Piglionica[3], A. De Luisi[3], F. C. Susca[3], N. Bukvic[3], C. Forleo[8], A. Selicorni[9], G. Zampino[2], G. Cappuccio[10], L. Garavelli[11], C. Novelli[12], L. Memo[13], C. Morando[13], M. Capurso[3], A. Cereda[14], M. Digiacomo[15], V. Saletti[16], A. M. Spinelli[17], P. Lastella[18], V. Dvorakova[19], A. D. Irvine[19], N. Resta[20]; [1]Pediatric Clinical Genetics Unit, Dept. of Publ. Hlth.and Pediatric Sci., Univ. of Torino, and Regina Margherita Childrenâ□□s Hosp., CittÃ della Salute e della Scienza, Torino, Italy, [2]Ctr. for Rare Diseases and Birth Defects, Dept. of Woman and Child Hlth.and Publ. Hlth., Fondazione Policlinico Univ.rio A. Gemelli IRCCS, Roma, Italy, [3]Dept. of BioMed. Sci. and Human Oncology (DIMO), Div. of Med. Genetics, Univ. of Bari â□□Aldo Moroâ□□, Bari, Italy, [4]Rare Diseases and Med. Genetics Unit, Bambino GesÃ¹ Children's Hosp. IRCCS, Roma, Italy, [5]Dept. of Clinical and Biological Sci., Univ. of Torino, Torino, Italy, [6]Dept. of Med., Surgery and Dentistry 'Scuola Medica Salernitana', Univ. of Salerno, Salerno, Italy, [7]Clinical Pediatric Genetics Unit, Pediatrics Clinics, MBBM Fndn., S. Gerardo Hosp., Monza, Italy, [8]Cardiology Unit, Dept. of Emergency and Organ Transplantation, Univ. of Bari Aldo Moro, Bari, Italy, [9]Centro Fondazione Mariani per il Bambino Fragile ASST-Lariana Sant'Anna Hosp., Dept. of Pediatrics, San Fermo della Battaglia, Como, Italy, [10]Dept. of Translational Med., Section of Paediatrics, Federico II Univ., Napoli, Italy, [11]Med. Genetics Unit, Mother and Child Hlth.Dept., Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy, [12]Dept. of Clinical Sci. and Community Hlth., The Univ. of Milan, Milano, Italy, [13]Dept. of Pediatrics, Neonatal Intensive Care Unit, San Bortolo Hosp., Vicenza, Italy, [14]Pediatric Dept., ASST Papa Giovanni XXIII, Bergamo, Italy, [15]Unit of Pathology and Med. Genetics, Azienda Ospedaliera Regionale San Carlo, Potenza, Italy, [16]Dept. of Pediatric NeuroSci., Fondazione IRCCS Istituto Neurologico Carlo Besta, Milano, Italy, [17]Regional Coordinating Ctr. for Rare Diseases, Univ. Hosp., Udine, Italy, [18]Centro Sovraziendale di Assistenza e Ricerca per le Malattie Rare, Internal Med. Unit 'C. Frugoni', Ospedale Consorziale Policlinico di Bari, Bari, Italy, [19]Dermatology, Children's Hlth.Ireland at Crumlin, Dublin, Ireland, [20]Dept. of BioMed. Sci. and Human Oncology (DIMO), Div. of Med. Genetics, Univ. of Bari â□□Aldo Moro, Bari, Italy

Disclosure Block: **D. Carli:** None.

Purpose: Post-zygotic activating *PIK3CA* variants are responsible for the several phenotypes under the *PIK3CA*-Related Overgrowth Spectrum (PROS). We describe 150 new patients referred for genetic testing, reporting data on their phenotype and underlined *PIK3CA* or *GNAQ, GNA11, RASA1*, and *TEK* variants. A detailed genotype-phenotype correlation in a literature-derived cohort of 1007 *PIK3CA*-mutated patients was also provided. **Methods:** We performed targeted NGS on DNA extracted from blood or buccal swab and tissue biopsy using a custom panel including genes involved in the PI3K/AKT/mTOR pathway and â□□vascularâ□□ genes (*RASA1, TEK, GNAQ,* and *GNA11*). Clinical and molecular data from literature were systematically reviewed and included in the search for correlations. **Results:** *PIK3CA* pathogenetic variants were identified in 93 of 150 unrelated patients. Fifty-seven had a wild type *PIK3CA* allele: pathogenetic variants in *GNA11*, *RASA1, GNAQ*, and *TEK* were found in 11 of them. Differences in the distribution of the variants across *PIK3CA* domains, in the Variant Allele Fraction (VAF) in different tissues, and in the degree of hyperactivation of the PI3K network related to the variant (variant strength) were found compared to current literature. Alone, 10 of the *PIK3CA* variants reported were responsible for more than 70% of cases, including the three most common mutational hotspots usually reported in several cancers. Combining our cases with those from literature review we draw up a detailed list of all the 81 pathogenic variants described so

far in PROS and related them to the respective phenotypes. Eight novel pathogenic variants were also reported. While some *PIK3CA* variants were exclusively associated with a specific PROS phenotype, some were scattered across all the phenotypes and others demonstrated enrichment in some specific phenotypes. Correlations between variant strength and absence of involvement of the central nervous system were also evident. VAF was not correlated with disease severity, and we report severe phenotypes with very low VAF. Patients with pathogenic variants in vascular genes clinically overlapped with PROS. **Conclusion:** Our findings combined with a review of the literature show novel genotype-phenotype correlations underlining the importance of performing a deep phenotyping, carry out a representative tissue sampling, and adopt a comprehensive molecular approach in PROS and overlapping phenotypes.

PrgmNr 2559 - Germline 3q21 Chromosome Deletion: A Rare Multi-System Condition

View session detail

Author Block: **J. Greenmyer**, W. S. Thompson, N. L. Hoppman, L. A. Schimmenti, M. A. Kohorst; Mayo Clinic, Rochester, MN
Disclosure Block: **J. Greenmyer:** None.
Background and aims: Germline 3q21 chromosome deletions result in a rare multi-system condition with diverse phenotype. The aim of this study was to conduct a systematic review of the clinical characteristics of 3q21 deletion in order to draw conclusions about the multiple aspects of care these patients should receive.**Methods:** We report our multidisciplinary observations of a patient with 3q21 deletion. Additionally, we conducted a systematic literature review for studies describing 3q21 deletion. Our systematic search strategy included Ovid MEDLINE, EBM Reviews, Embase, Scopus, and Web of Science. Any deletion including the germline 3q21 region was included in our analysis.**Results:** We identified 13 manuscripts that described 14 unique cases of 3q21 region deletion. The deleted interval in our patient (3q21.3-3q22.1) involves 77 known genes, some of which are recognized to result in abnormal phenotypes due to haploinsufficiency. Clinical features among patients with 3q21 deletion include cognitive and developmental disability, hearing loss, cardiac abnormalities, facial differences (dysmorphism), velopalatal insufficiency, retinal lacunae, myelodysplastic syndrome, urogenital malformations, and other birth defects. Agenesis of the corpus callosum was noted in four (28.6%, n=4/14) patients. Two patients (14.3%, n=2/14) developed myelodysplastic syndrome.**Conclusions**: Patients with 3q21 deletion have a predilection for multi-system complications including agenesis of the corpus callosum.

PrgmNr 2560 - Neuropsychiatric and behavourial phenotype description of 24 Brazilian patients harboring Cornelia de Lange Syndrome

View session detail

Author Block: **L. Pires**[1], M. Rivadeneira[1], J. Ceroni[1], J. V. dos Santos[1], V. H. E. Kim[1], I. Furquim[1], H. Aoi[2], N. Matsumoto[2], C. B. Mello[3], D. M. Befi-Lopes[4], R. S. Honjo[1], D. R. Bertola[1], C. A. Kim[1]; [1]Genetic Unit - Inst. da CrianÃ§a do Hosp. das Clinicas da Univ.e de SÃ£o Paulo, SÃ£o Paulo, Brazil, [2]2.Dept. of Human Genetics, Yokohama City Univ., Yokohama, Japan, [3]Departamento de Psicobiologia, Univ.e Federal de SÃ£o Paulo, SÃ£o Paulo, Brazil, [4]Departamento de Fisioterapia Fonoaudiologia e Terapia Ocupacional, Faculdade de Med. da Univ.e de SÃ£o Paulo, SÃ£o Paulo, Brazil
Disclosure Block: **L. Pires:** None.
Introduction: Cornelia de Lange syndrome (CdLS) is a rare genetic disorder caused by mutations in six different genes - *NIPBL, SMC1A, HDAC8, SMC3, BRD4 and RAD21.* characterized by intellectual disability, behavior abnormalities, distinctive facial features, hirsutism, growth retardation and limb reduction defects.**Aim:** To investigate the Neuropsychiatric and Behavioral phenotype of 24 brazilian CdLS patients.**Methods**: We evaluated the cognitive and neuropsychiatric phenotype of 24 Brazilian patients harboring CdLS clinical phenotype and pathogenic variants on genes associated with the syndrome, diagnosed by exome sequencing of the proband, father, and mother.**Results:** Our sample was composed of 24 individuals, 15 male and 9 female, age of diagnosis ranged from 1 month-old to 9 years (mean 13 months; median 4 months). The actual age of the patients ranged from 4 years to 43 years (mean 18 years; median 16 years). The major genetic etiology was *NIPBL* (21 patients), followed by *SMC1A* (3 patients); all pathogenic variants were *de novo*. Neurodevelopmental delay was a major found on the sample, present in all patients. Although 17 patients (70,8%) attempted to go to school, only two were literate, six recognized alphabet letters and seven were able to recognize numbers. Feeding difficulties in the first year of life and Inattention problems were found in 18 patients (75%). Intolerance to pain was present in 16 patients (66%). Excessive fear and hyperactivity were both present in 13 patients (54,1%). Aggressive behaviour was present in 10 patients (41,6%), besides self-violence was present in 12 (50%). Anxiety was present in 10 patients (41,6%). Autistic Spectrum Disorder was present in 6 patients (25%). **Conclusions:** Neuropsychiatric symptoms and behavioral deviations are an important group of findings in CdLS. Due to high incidence found on psychiatric disorders, we suggest that neuropsychological evaluation and directive therapies should be a primal focus during diagnosis and treatment. Another important approach in these patients involves their learning disability, assuming that they might need special support to ensure a minimum acquisition of learning skills in school.

PrgmNr 2561 - Sequencing analysis of genes involved in the Wnt signaling pathways reveals novel candidates for nonsyndromic cleft lip with or without cleft palate

View session detail

Author Block: **L. A. Brito**[1], G. S. Kobayashi[1], G. S. Hsia[1], D. P. Moreira[1], J. Y. T. Wang[1], S. G. Ferreira[1], L. Alvizi[1], G. L. Yamamoto[2], M. R. Passos-Bueno[3]; [1]Inst. of BioSci.s, Univ. of SÃ£o Paulo, SÃ£o Paulo, Brazil, [2]Inst. da CrianÃ§a - Univ. of SÃ£o Paulo, SÃ£o Paulo, Brazil, [3]Inst. of BioSci.s, Univ. of SÃ£o Paulo, Sao Paulo, Brazil

Disclosure Block: **L.A. Brito:** None.

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a complex disease, for which genetic and environmental factors play etiological roles. Sequencing and functional approaches have demonstrated the contribution of common and rare variants to NSCL/P, along with epigenetic signatures. Many NSCL/P-associated genes converge to the Wnt signaling pathways or interacting partners. In this study, we investigated the presence of rare variants in a custom panel of 27 genes directly or indirectly involved with the canonical Wnt / beta catenin and the noncanonical Wnt / Planar Cell Polarity (PCP) pathways: *ARHGAP29, CBLL1, CDH1, CELSR3, CTNNA1, CTNNB1, CTNND1, DVL3, ESRP1, ESRP2, FZD1, FZD2, FZD7, LRP6, PRICKLE1, PRICKLE2, RHOA, SMAD2, SMAD4, SNAIL, TGFB1, VANGL2, WNT4A, WNT10A, WNT11, YAP1, ZEB1* and *ZEB2*. After sequencing with the Illumina Ampliseq platform in MiSeq system, 309 patients, all from familial cases of NSCL/P, presented >90% of bases with 20x minimum coverage and proceeded to subsequent variant analysis. Rare (20 were filtered. Compared to ethnically matched controls from ABraOM database (exome sequencing data; only segments with 20x minimum coverage were compared), no aggregation of such variants in any gene was observed in patients (P >=0.07). Among all prioritized variants, criteria for variant classification from the American College of Medical Genetics support a pathogenic role for 19 variants, mostly located in NSCL/P-associated genes - *CTNND1* (6), *CDH1* (5) and *ARHGAP29* (2) -, but some of them in novel candidates: *CTNNA1* (3), *ESRP1* (1), *PRICKLE1* (1) and *ZEB1* (1). Segregation analysis, which was possible for only 5 families, revealed complete segregation of variants in *CDH1* (2), *CTNND1, CTNNA1* and *PRICKLE1* with NSCL/P, corroborating their pathogenic effect. *CDH1*, *CTNNA1* and *CTNND1* codify key proteins of the cadherin-catenin complex of adhesion junctions in epithelial cells, which dynamics is regulated by canonical Wnt pathway during embryogenesis. Interestingly, the variant segregating in *CTNNA1* is a missense one located at the subdomain VH1, responsible for the attachment to beta-catenin in the adhesion complex. Moreover, this family reports cases of gastric cancer, a phenotype that has been associated with *CDH1* and CTNNA1. In conclusion, we report novel pathogenic variants in *CDH1*, *CTNND1* and *ARHGAP29* and suggest the pathogenic effect of rare variants in novel candidate genes, such as *CTNNA1*, *ESRP1*, *PRICKLE1* and *ZEB1*. Functional analyses will be of interest to confirm the effect of the variants here reported

PrgmNr 2562 - A homozygous missense variant in amyloid-beta precursor protein *(APP)* may cause severe syndromic intellectual disability

View session detail

Author Block: **K. Riquin**[1], T. Besnard[1,2], A. Toutain[3], V. Vignard[1,2], S. Bezieau[1,2], B. CognÃ©[1,2]; [1]UniversitÃ© de Nantes, CHU Nantes, CNRS, INSERM, lâinstitut du thorax, Nantes, France, [2]Ctr. Hosp.ier Univ.ire de Nantes, Service de GÃ©nÃ©tique MÃ©dicale, Nantes, France, [3]Service de GÃ©nÃ©tique, Ctr. Hosp.ier Univ.ire de Tours, Inserm U1253, Ibrain, UniversitÃ© de Tours, Tours, France

Disclosure Block: **K. Riquin:** None.

Intellectual disability is a clinically and genetically heterogeneous group of disorders. It is characterized by deficits in intellectual and adaptive functioning begining before adulthood, with a lasting effect on development. Recently, we identified by genome sequencing a homozygous missense variant NM_201414.3:c.440A>G:p.(His147Arg) in the E1 domain of Amyloid-beta precursor protein *(APP)* in one patient with severe intellectual disability, absent speech, epilepsy, cerebellar hyploplasia, and autistic behaviour. APP is an ubiquitous cell surface receptor able to homo- or hetero-dimerize. The most studied isoform of APP is the 695 amino-acids protein which is expressed at the surface of neurons. It has been observed that APP dimerization at neuron membrane participate to neurite growth, neuronal adhesion, axonogenesis and synaptogenesis. Many pathogenic missense variants have already been identified in the AÃ domain and are associated with the accumulation of amyloid Ã peptide leading to early onset alzheimer disease. To date, only one homozygous loss-of-function variant in *APP* was reported, in an individual presenting with global developmental delay, epilepsy and microcephaly (Klein et al. 2016). The missense variant we identified substitutes a cupper-binding histine that is required for APP dimerization. We are currently trying to modelize the functional effect of the variant in HEK293 and SH-SY5Y cell lines. We will present the first results of this functional study and we will discuss the implication of *APP* biallelic loss-of-function variants in syndromic intellectual disability.

PrgmNr 2563 - A nonsense *EBF3* variant in an individual with intellectual disability, without ataxia and hypotonia

View session detail

Author Block: **S. Spineli-Silva**[1], N. de Leeuw[2], N. Leijsten[2], M. H. A. Ruiterkamp-Versteeg[2], J. R. M. Prota[1], A. P. Marques-de-Faria[1], T. P. Vieira[1]; [1]Dept. of Translational Med., Sch. of Med. Sci., State Univ. of Campinas, Campinas, Brazil, [2]Dept. of Human Genetics, Radboud Univ. Med. Ctr., Nijmegen, Netherlands

Disclosure Block: **S. Spineli-Silva:** None.

Heterozygous variants in the Early B cell factor 3 (*EBF3*) and deletions at the 10q26.3 region, encompassing the entire *EBF3* gene, have been reported in individuals presenting with hypotonia, ataxia and delayed development syndrome (HADDS) (MIM#617330), first described in 2017. However, individuals with pathogenic variants in the *EBF3* show phenotypic heterogeneity. We report on a heterozygous *de novo* variant in the *EBF3* gene in an individual with neurodevelopmental delay, behavioural manifestations, and craniofacial features. The proband, an eleven years-old boy, the only child of a non-consanguineous and healthy couple, was referred for genetic evaluation due to speech delay, learning disability and behavioural problems that suggest an oppositional defiant disorder. It was also related hyperactivity, irritability, hetero-aggressiveness, visual hallucinations, insomnia and decreased pain sensitivity. At physical evaluation, it was observed macrocephaly, long face, prominent nasal bridge, everted lower lip, high palate, clinodactyly of 5th finger, hypoplasia of 4th and 5th metacarpus, and pes planus. The chromosomal microarray analysis showed a gain of 365 kb in Xp22.33 or Yp11.32 (pseudo autosome region 1) that was classified as a Variant of Uncertain Significance. Subsequent whole exome sequencing, performed with the Agilent SureSelect Target Enrichment V5 (Agilent TechnologiesÂ®) and the Illumina HiSeq platform (Illumina$^{Â©}$, Inc.), revealed a heterozygous nonsense variant c.1381C>T (p.Arg461*) in the *EBF3* gene (NM_001005463.2), classified as pathogenic. Validation of this variant and segregation analysis in the parents were performed by Sanger sequencing and confirmed this variant to be *de novo*. This truncating variant in the *EBF3* gene (c.1381C>T) has been reported in a single individual in the gnomAD Database. However, pathogenicity classification and clinical data are not available. Recently, this gene has been included in the Gene Reviews as the â□□*EBF3* Neurodevelopmental Disorderâ□□ and, to date, less than 50 cases have been described. The case herein described, without ataxia and hypotonia, supports the clinical variability of this condition.

PrgmNr 2564 - A predictive modeling strategy to explore the mechanism(s) underlying dominantly-inherited peripheral neuropathy caused by aminoacyl-tRNA synthetase variants

View session detail

Author Block: **A. Cale**, A. Antonellis; Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **A. Cale:** None.
Aminoacyl-tRNA synthetases (ARSs) are ubiquitously expressed and essential enzymes that ligate amino acids to the appropriate tRNA molecules in the cytoplasm and mitochondria; ARSs function as monomers or dimers. Mutations in five ARSs cause autosomal dominant, axonal peripheral neuropathy (aka Charcot-Marie-Tooth [CMT] disease), which is characterized by distal muscle weakness and sensory loss. This presents the question: how do mutations in ARSs, which are essential in all tissues, lead to tissue-specific effects? There are three possibilities for the molecular pathology: haploinsufficiency, dominant-negative effects, and gain-of-function effects. Haploinsufficiency has been ruled out based on genetic studies in mouse and human, and on the observation that pathogenic variants are limited to missense and small in-frame deletion variants. Interestingly, the five ARSs associated with CMT all function as homodimers, which is consistent with a dominant-negative effect. If a dominant-negative effect is the disease mechanism, it would be expected that certain variants in any cytoplasmic, dimeric ARS could lead to peripheral neuropathy. To test this, I am employing a predictive modeling strategy in which I engineer mutations in threonyl-tRNA synthetase (*TARS1*), which is a cytoplasmic dimeric ARS that has not been implicated in CMT. Based on the observation that a small in-frame deletion in the glycine binding pocket of glycyl-tRNA synthetase (*GARS1*) causes a severe, early-onset neuropathy, I engineered six human *TARS1* variants that similarly impact the threonine binding pocket. I developed a humanized *TARS1* yeast complementation assay, which revealed that all six of the engineered variants cause a loss-of-function effect, similar to the majority of CMT-associated ARS variants. I also performed western blot analysis, which revealed that two of the loss-of-function variants do not reduce protein expression; this last set of variants are the most promising to assess for a dominant-negative effect. Moving forward, I will: (i) test the mutations for dominant-toxicity in a yeast assay expressing wild-type and mutant *TARS1*; (ii) develop heterozygous worms for any dominantly toxic variants and assess for motor behavior deficits; and (iii) develop heterozygous mice for any promising variants and assess for peripheral neuropathy. Here, I will present all of our unpublished data. Successful completion of this study will provide insight into the mechanism underlying ARS-associated CMT and inform clinicians to screen all genes encoding a cytoplasmic dimeric ARS for pathogenic variants in patients with peripheral neuropathy.

PrgmNr 2565 - Biallelic frameshift variant in the *TBC1D2B* gene in two siblings with progressive cognitive impairment, gingival overgrowth, limb tremor, and fribous dysplasia of face

View session detail

Author Block: **G. ROLDÃ□O CORREIA COSTA**[1], N. de Leeuw[2], R. Pfundt[2], I. Cristina Sgardioli[1], A. dos Santos[1], V. Gil-da-Silva-Lopes[1], T. Paiva Vieira[1]; [1]Sch. of Med. Sci., Univ. of Campinas, Campinas, Brazil, [2]Radboud Univ. Med. Ctr., Nijmegen, Netherlands
Disclosure Block: **G. RoldÃ£o correia costa:** None.
The *TBC1D2B* gene is a GTPase-activating protein involved in membrane trafficking that interacts with the early endosomal marker proteins RAB5. Biallelic lossâ□□ofâ□□function variants in this gene were first reported in 2020 as a cause for a neurodevelopmental disorder with seizures and gingival overgrowth in individuals from three unrelated families. Here we report two male siblings, with similar clinical characteristics, born to a first-degree cousin couple. The oldest sibling started with bilateral growing of soft tissues in the malar region at three years old, which evolved with significant maxillary hypertrophy and compression of the brainstem. At 17 years old, he presented mental deterioration, limb tremors, ataxia, gingival overgrowth, and fibrous dysplasia. At his last evaluation, at 38 years old, he was bedridden and dependent on assisted ventilation. His younger brother presented with a similar clinical evolution, starting also at three years old. His condition evolved with the same characteristics and, at 27 years old, he was also bedridden and with tracheostomy. Chromosomal Microarray Analysis for both siblings did not show pathogenic CNVs, however, it showed multiple regions of homozygosity (ROH) in the autosomal genome of both. Whole Exome Sequencing was performed using the Agilent SureSelect Human All Exon V5 capture kit (Agilent Technologies$^{Â®}$) and the Illumina HiSeqâ□¢ 2000 platform (Illumina$^{Â©}$, Inc.). Data analyses, including annotation and variant classification, were carried at the Genomic Diagnostics Division from the Department of Human Genetics at the Radboud University Medical Center in Nijmegen, Netherlands. A novel biallelic frameshift variant in the *TBC1D2B* gene (NM_144572.1) was found in both siblings - Chr15(GRCh37): g.78337330del c.595del p.(Val199Trpfs*22) - which creates a new stop codon at position 22. This gene is encompassed in one of the ROHs shared by the two brothers. The homozygous variant was confirmed by Sanger sequencing in both siblings and was found in heterozygous form in each of their parents. There are strong similarities of clinical characteristics, and its evolution, among the patients described here and the reported cases, including a cherubism-like phenotype with progressive gingival overgrow. This is the fourth family in the world in which a bi-allelic loss-of-function variant in the *TBC1D2B* gene has been found, in two patients with similar phenotypes. These results support that loss of *TBC1D2B* is the cause for this rare condition.

PrgmNr 2566 - Biallelic SEPSECS variants in two siblings with pontocerebellar hypoplasia type 2D underscore the relevance of splice-disrupting synonymous variants in disease

View session detail

Author Block: **S. Ramadesikan**[1], S. E. Hickey[2,3], E. Crist[1], S. Franklin[1], P. Brennan[1], P. White[4,3], K. L. McBride[5,2,3], D. C. Koboldt[5,3], R. K. Wilson[6,3]; [1]The Steve and Cindy Rasmussen Inst. for Genomic Med., Nationwide Children's Hosp., Columbus, OH, [2]Div. of Genetic and Genomic Med., Nationwide Children's Hosp., Columbus, OH, [3]Dept. of Pediatrics, The Ohio State Univ. Coll. of Med., Columbus, OH, [4]Nationwide Children's Hosp., Columbus, OH, [5]Nationwide Children's Hosp, Columbus, OH, [6]Nationwide Childrenâ□□s Hosp. Inst. for Genomic Med., Columbus, OH

Disclosure Block: **S. Ramadesikan:** None.

Noncoding and synonymous coding variants that exert their effects via alternative splicing are increasingly recognized as an important category of disease-causing variants. In this study, we describe two siblings, a 16-year-old female and 9-year-old male, who presented with hypotonia at birth, profound developmental delays, and seizures. MRI brain in the male at 5 years showed diffuse cerebral and cerebellar white matter volume loss. The siblings later developed ventilator-dependent respiratory insufficiency, scoliosis and are currently nonverbal and non-ambulatory. Extensive molecular testing including oligo array and clinical exome sequencing was non-diagnostic. Research genome sequencing under an IRB-approved study protocol revealed that both affected children were compound-heterozygous for variants in the *SEPSECS* gene. One variant was an initiator codon change (c.1A>T) that disrupted protein translation, consistent with the observation that most disease-causing variants are loss-of-function changes. The other variant, however, was a coding change (c.846G>A) that was predicted to be synonymous but had been demonstrated to disrupt mRNA splicing in a minigene assay.

SEPSECS gene encodes O-phosphoseryl-tRNA(Sec) selenium transferase which participates in the biosynthesis and transport of selenoproteins in the body. Mutations in *SEPSECS* cause autosomal recessive pontocerebellar hypoplasia type 2D (PCHT 2D; OMIM #613811), a neurodegenerative condition characterized by progressive cerebrocerebellar atrophy, microcephaly, and epileptic encephalopathy. The identification of biallelic pathogenic variants in this family - one of which was a synonymous change not reported in prior clinical testing - not only ended the diagnostic odyssey for this family, but also highlighted the contribution of occult pathogenic variants that may not be recognized by standard genetic testing methodologies. Further, studies like this highlight the importance and relevance of genome sequencing in the diagnosis of undiagnosed pediatric conditions.

PrgmNr 2567 - CAPRIN1: how a recurrent *de novo* mutation causes a progressive early onset neurodegenerative disorder in a prion-like domain-harboring protein

View session detail

Author Block: **A. Delle Vedove**[1,2,3], J. GuillÃ©n Boixet[4], G. Zanni[5], M. Eckenweiler[6], A. MuiÃ±os BÃ¼hl[1,2,3], M. Storbeck[1,2,3], S. Barresi[5], S. Pizzi[5], I. HÃ¶lker[1,2,3], F. KÃ¶rber[7], E. Bertini[5], J. Kirschner[8,6], S. Alberti[4], M. Tartaglia[5], B. Wirth[1,2,3]; [1]Inst. of Human Genetics, Univ. of Cologne, Cologne, Germany, [2]Ctr. for Molecular Med. Cologne, Univ. of Cologne, Cologne, Germany, [3]Inst. for Genetics, Univ. of Cologne, Cologne, Germany, [4]Ctr. for Molecular and Cellular Bioengineering, Biotechnology Ctr., Technische Univ. Dresden, Dresden, Germany, [5]Genetics and Rare Diseases Res. Div., Ospedale Pediatrico Bambino GesÃ¹, Rome, Italy, [6]Dept. of Neuropediatrics and Muscle Disorders, Med. Ctr. â Univ. of Freiburg, Faculty of Med., Univ. of Freiburg, Freiburg, Germany, [7]Inst. of Diagnostic and Interventional Radiology, Cologne, Germany, [8]Dept. of Neuropediatrics, Univ. Hosp. Bonn, Bonn, Germany
Disclosure Block: **A. Delle Vedove:** None.

CAPRIN1 is a ubiquitously expressed protein, abundant in the brain, where it regulates the transport and translation of mRNAs of genes involved in synaptic plasticity. Like other proteins (TDP-43, FUS or TIA1) related with neurodegenerative disorders (ND), CAPRIN1 is a component of stress granules and harbours a prion-like domain.

Here we describe two unrelated children from non-consanguineous families of Turkish and Italian ancestry, who respectively developed at 10 years and 7 years of age ataxia and cognitive decline. Trio whole exome sequencing unravelled an identical *de novo* c.1535C>T variant (p.Pro512Leu) in *CAPRIN1*. This variant is not reported in gnomAD and affects a highly conserved residue. Applying *in silico* prediction tools (PLAAC, Zyggregator, Aggrescan), we found an increased aggregation propensity of the p.Pro512Leu mutated protein, suggesting that protein or protein-RNA aggregation might be the underlying pathomechanism for this ND. Indeed, we observed that overexpression of the mutated but not wild type CAPRIN1 caused the formation of insoluble aggregates in transfected HEK293T and SH-SY5Y cells, as shown by immunostaining analysis and confirmed by biochemical studies. Strikingly, this change in solubility depends on the CAPRIN1 interaction with RNA: *in vitro* mCherry-CAPRIN1 remains soluble, while a fraction of mCherry-CAPRIN1^{P512L} aggregates. Moreover, we generated CAPRIN1^{P512L} heterozygous and homozygous hiPSC lines using the CRISPR/Cas9 system and differentiated them into cortical neurons, which show reduced axonal growth and reduced neuronal activity.

In conclusion, we identify a missense variant in *CAPRIN1* as a crucial change associated with an early onset neurodegenerative disorder, in contrast to *CAPRIN1* haploinsufficiency that is related to autism-spectrum disorders. CAPRIN1^{P512L} is linked to RNA-dependent increased aggregation propensity and morphological and electrophysiological alterations in neurons.

PrgmNr 2568 - Characterization of Phelan-McDermid brazilian cohort with small deletions and point mutations: focus on regression and behavioral changes across different ages

View session detail

Author Block: **E. Varella Branco**[1], E. Casella[2], I. Ramos[1], C. Samogy Costa[1], A. C. Girardi[1], H. Ferraz[3], M. R. Passos-Bueno[4]; [1]Inst. de BiociÃªncias, Univ.e de SÃ£o Paulo, SÃ£o Paulo, Brazil, [2]Unidade de Neuropediatria do Inst. da CrianÃ§a, Hosp. da ClÃnicas da Faculdade de Med., Univ.e de SÃ£o Paulo, SÃ£o Paulo, Brazil, SÃ£o Paulo, Brazil, [3]Programa de Engenharia QuÃmica, Univ.e Federal do Rio de Janeiro, Rio de Janeiro, Brazil, [4]Univ De Sao Paulo, Sao Paulo, Brazil
Disclosure Block: **E. Varella Branco:** None.

Phelan-McDermid Syndrome (PMS) is a rare genetic disorder mainly characterized by global developmental delay, intellectual disability (ID), autism spectrum disorder (ASD) caused by *SHANK3* loss-of-function mutations. Deletions (few kb to more than 9 Mb) at 22q13.3 are the most common mutational mechanisms, but point disruptive variants in *SHANK3* accounts for about 3% of the cases. Even though neurological phenotypes, like ASD and ID has been extensive described, changes in behavior (catatonia, schizophrenia) or regressions (loss of acquired abilities) during growth of these patients is still poorly characterized. It is still unclear if these changes are dependent on haploinsufficiency of *SHANK3* or other extrinsic factors. To address additional insights into these questions and to better understand the long-term course of the disorder, we aim to delineate the associated regression and behavior changes, in a group of PMS patients with small deletions or point mutations. We evaluated a total of 13 patients: nine patients with small deletions in 22q13.3 (49kb to 110kb), and four patients with pathogenic point mutations in *SHANK3*. Eight patients currently at 4 to 39 years old have developed speech until 4 years, however seven of them lose this ability. A history of regression, affecting primarily language ability and social ability at any age was present in 92% (12/13) of all patients. The onset of regression has been reported in different age periods: 11 months (n=1), 2-4 years old (n= 7), 6-7 years old (n= 2), and after 16 years old (n=2). Especially in cases of late regression, patients showed signs of behavioral and neurologic decompensation, with clinical presentations of catatonia, psychosis and schizophrenia. Our work corroborates the results described by others related to regression in PMS, which show most frequently episodes have an onset in early to middle-childhood, and highlighted the behavior changes in long-term course of the disorder in patients with *SHANK3* haploinsufficiency. These results may suggest that *SHANK3* is important for the maintenance of the nervous system in more advanced stages of neurodevelopment, which so far have not been elucidated. Furthermore, confirmation of similar findings in an ethnically different population, reinforce that the mechanisms underlying regression is mainly genetic dependent. Finally, the identifications of biological mechanisms leading to regressions in PMS are extremely important to improve interventions along life. This work is being perfomed in collaboration with PMS Brazilian Association (Amigos e Familiares da SÃndrome de Phelan-McDermid - AFSPM). Financial support: CEPID/ FAPESP.

PrgmNr 2569 - Characterization of the GARS1 mRNA interactome in glycyl-tRNA synthetase (*GARS1*)-related neurological disease

View session detail

Author Block: **M. Forrest**[1], O. Levi[2], Y. Arava[2], A. Antonellis[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Technion-Israel Inst. of Technology, Haifa, Israel
Disclosure Block: **M. Forrest:** None.
Aminoacyl-tRNA synthetases (ARSs) are a group of ubiquitously expressed essential genes that conjugate amino acids to cognate tRNAs in the cytoplasm and mitochondria, a critical process in protein synthesis and translation fidelity. Intriguingly, the phenotypic spectrum of ARS-associated disease ranges from early-onset, multi-system, recessive phenotypes with significant neuronal involvement to late-onset dominant peripheral neuropathies. Notably, heterozygosity for loss-of-function missense mutations in the bifunctional glycyl-tRNA synthetase (*GARS1*) gene cause either axonal Charcot-Marie-Tooth disease (CMT2) or infantile-onset spinal muscular atrophy (iSMA). However, the molecular mechanisms underlying these nervous system-specific disease phenotypes remain poorly understood.
One explanation for the predominance of neurological phenotypes associated with GARS1-related disease is that GARS1 may serve functions outside of tRNA charging in neurons. Indeed, recent studies have demonstrated non-canonical mRNA binding by various ARSs in yeast and human cells, suggesting a role for the tRNA anticodon binding domain (ABD) in mRNA target recognition and post-transcriptional gene regulation. We hypothesize that GARS1 binds and regulates a subset of mRNAs that are involved in nervous system function or maintenance, with potential relevance to *GARS1*-related neurological disease. To characterize the GARS1-mRNA interactome in human cells, we used RNA immunoprecipitation coupled to deep RNA sequencing (RIP-seq) to identify significantly enriched mRNAs bound by wild-type GARS1 in human cells. Preliminary results demonstrate significant enrichment of 35 RNA transcripts (log2-fold change over input control >2, p-value

PrgmNr 2570 - Complex inheritance in spastic paraplegia 7

View session detail

Author Block: **M. A. Estiar**[1,2], E. Yu[1], I. Haj Salem[3], J. P. Ross[1], K. Mufti[1], F. AkÃ§imen[1], E. Leveille[4], D. Spiegelman[2], J. Ruskey[2], F. Asayesh[2], A. Dagher[5], G. Yoon[6], M. Tarnopolsky[7], K. Boycott[8], N. Dupre[9,10], P. Dion[11], O. Suchowersky[12], J-F. Trempe[13,14], G. Rouleau[2,11,1], Z. Gan-Or[11,2,1]; [1]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, [2]The Neuro (Montreal Neurological Inst.-Hosp.), McGill Univ., Montreal, QC, Canada, [3]FacultÃ© de MÃ©decine, UniversitÃ© Laval, QuÃ©bec City, QC, Canada, [4]Faculty of Med., McGill Univ., Montreal, QC, Canada, [5]The Montreal Neurological Inst.-Hosp.), McGill Univ., Montreal, QC, Canada, [6]Div.s of Neurology and Clinical and Metabolic Genetics, Dept. of Paediatrics, Univ. of Toronto, Toronto, ON, Canada, [7]Div. of Neuromuscular and Neurometabolic Disorders, Dept. of Pediatrics, McMaster Univ., Hamilton, ON, Canada, [8]Childrenâ□□s Hosp. of Eastern Ontario Res. Inst., Univ. of Ottawa, Ottawa, ON, Canada, [9]NeuroSci. Axis, CHU de QuÃ©bec, UniversitÃ© Laval, QuÃ©bec City, QC, Canada, [10]Dept. of Med., Faculty of Med., UniversitÃ© Laval, QuÃ©bec City, QC, Canada, [11]Dept. of Neurology and Neurosurgery, McGill Univ., Montreal, QC, Canada, [12]Dept.s of Med. (Neurology) and Med. Genetics, Univ. of Alberta, Edmonton, AB, Canada, [13]Dept. of Pharmacology & Therapeutics, McGill Univ., Montreal, QC, Canada, [14]Ctr. de Recherche en Biologie Structurale, McGill Univ., Montreal, QC, Canada

Disclosure Block: **M. A. Estiar:** None.

Hereditary spastic paraplegia (HSP) is often considered a monogenic Mendelian disorder including all modes of inheritances. According to some reports, spastic paraplegia type 7 (SPG7) is the most common form of the autosomal recessive HSP, while an autosomal dominant form has also been suggested. In this study, we aimed to conduct a genetic analysis on one of the largest cohorts of HSP to find the spectrum of genetic inheritance of SPG7. In total, 585 HSP patients (372 probands) and 1175 controls (580 unrelated individuals) were analyzed using whole-exome sequencing and sequencing a targeted panel of genes. Among unrelated patients (N=291) and controls (N=580) who went through whole-exome sequencing, the number of heterozygous pathogenic and likely pathogenic variants of *SPG7* in patients was higher compared to controls (OR 2.88, 95% CI 1.24-6.66, P=0.009). The frequency of common heterozygous pathogenic *SPG7* variant, p.(Ala510Val), was 3.7% in patients vs. 0.85% in controls (OR 4.42, 95% CI 1.49-13.07, P=0.005). We found 4 patients carrying a heterozygous *SPG7* variant along with a pathogenic variant in another phenotype-associated gene (*SPAST*, *BSCL2* and *TBCE*). We further examined variants in genes that produce proteins that interact with SPG7, and found heterozygous variants in three genes (*CACNA1A*, *AFG3L2*, and *MORC2*). Of these, there is additional functional and protein structural evidence for co-occurrence of the variants of *SPG7* and *AFG3L2* genes. Our findings provide evidence for non-Mendelian inheritance in SPG7-HSP, including autosomal recessive, possibly dominant and digenic modes of inheritance.

PrgmNr 2571 - Defining the molecular mechanism of*GARS1*-related Charcot-Marie-Tooth disease

View session detail

Author Block: **S. Marte**, A. Antonellis; Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **S. Marte:** None.
Aminoacyl-tRNA synthetases (ARSs) are essential enzymes required to charge tRNA molecules to cognate amino acids in the cytoplasm and mitochondria. Although ARSs are essential and ubiquitously expressed, loss-of-function (LOF) missense mutations in five dimeric ARS enzymes have been associated with dominant peripheral neuropathy (also known as Charcot-Marie-Tooth disease [CMT]). CMT is a genetically and clinically heterogeneous inherited peripheral neuropathy characterized by the progressive loss of motor and sensory function. Mutations in glycyl-tRNA synthetase (*GARS1*) have been associated with distinct clinical phenotypes where individuals can present with later-onset CMT or infantile spinal muscular atrophy. The mechanism by which mutations in *GARS1* lead to distinct clinical phenotypes is currently unclear. Since all five implicated ARSs function as dimers and since CMT-associated ARS variants all cause a loss-of-function effect, we propose the possibility of a dominant-negative mechanism. To test dominant-negative effects of pathogenic glycyl-tRNA synthetase (*GARS1*) variants, we will develop a humanized yeast model and test *GARS1* mutations for the ability to repress a wild-type copy of *GARS1*. To better understand the distinct clinical phenotypes, we will assess the dominant toxicity of a series of pathogenic *GARS1* alleles to determine if toxicity in our yeast model correlates with disease severity. Finally, we will identify pathways that, when manipulated, improve *GARS1* function by performing experimental evolution and gain-of-function studies using a hypomorphic allele and yeast growth assays. Here, we will present preliminary data on development of our humanized model and on initial assessments of the dominant toxicity of pathogenic *GARS1* alleles. These studies will provide insight into the pathogenic mechanism of ARS-related CMT and aid in the development of therapeutics.

PrgmNr 2572 - Defining the neurodevelopmental spectrum of *GABRA5*-associated disease

View session detail

Author Block: **R. Olson**[1], K. Johannesen[2], K. Grand[3], J. M. Graham[3], W. Mu[4], K. Baranano[4], S. Andrews[5], L. Faivre[6], C. Philippe[6], V. Desportes[7], B. Lanpher[1], R. MÃ¸ller[2], K. Eric[1], P. Ahring[8]; [1]Mayo Clinic, Rochester, MN, [2]Danish Epilepsy Ctr., Dianalund, Denmark, [3]Cedars-Sinai Med. Ctr., Los Angeles, CA, [4]Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [5]Hosp. for Sick Kids Toronto, Toronto, ON, Canada, [6]UniversitÃ© Bourgogne-Franche ComtÃ©, Dijon, France, [7]HÃ´pital Femme MÃ¨re Enfant, Bron, France, [8]The Univ. of Sydney, New South Wales, Australia

Disclosure Block: **R. Olson:** None.

The GABA-A receptors are Pentameric Ligand-Gated Ion Channels responsible for mediating inhibitory neurotransmission. The GABA-A receptor typically includes two alpha, two beta and one gamma subunit, with several genes coding for each subunit. Pathogenic variants disrupting the GABAergic system have been associated with a wide spectrum of neurodevelopmental disorders. *GABRA5*, one of a six alpha genes, has been recently associated with Developmental and Epileptic Encephalopathy (DEE) 79 (MIM #618559), with only seven variants reported in the literature. These private or rare variants occurred de novo or were inherited from unaffected parents, suggesting a partial penetrant phenotype, with functional testing reported to be loss-of-function of GABA-A receptor activity.

Here, we describe six additional patients (3-18yo) found to carry a *GABRA5* heterozygous variant. One individual inherited the variant from an affected mother. All but one of the de novo cases had a severe phenotype with early DEE and an EEG revealing rapid activity consistent with diffuse disturbance of cerebral function and largely absent sleep architecture. The remaining individual had similar EEG findings, but to date no seizures. Also, these individuals present with moderate to severe motor and cognitive delays and delayed to absent speech. The individual with the rare inherited variant is diagnosed with epilepsy and similarly, is developmentally delayed and non-verbal. Additional disease presentations were inconsistent, with three patients having dysmorphic features and one patient experiencing generalized hypotonia and joint hyperlaxity. Consistent with two previously published variants, p.(Val294Leu) and p.(Val294Phe), we identified two additional variants occurring in the pore of the channel, p.(Val291Leu) and p.(Thr301Arg), indicating that disruption of the ion selectivity filter is disease causing. Two additional variants, p.(Asn342His) and p.(Tyr225Asn), identified in our cohort, respectively occur in the M3-M4 intracellular or the extracellular domains. In silico analysis of each variant are highly suggestive of a deleterious effect on GABRA5 function (REVEL>0.7, CADD>20 and PolyPhen2>0.995).

In summary, we further delineate the phenotypic spectrum of *GABRA5*-associated disease, which will aid genetic diagnosis and counseling in these individuals and help understand the disease course. In addition, functional analysis of these and previously published variants are being pursued to establish pathogenicity and disease mechanism with the goal to better understand how GABRA5 dysregulation leads to seizure activity.

PrgmNr 2573 - Elucidating genomic and transcriptomic hexanucleotide repeat number and methylation in the X-linked dystonia-parkinsonism-relevant retrotransposon insertion by Nanopore sequencing

View session detail

Author Block: **J. LaÃ**[1], S. Schaake[1], I. Wohlers[2,3], T. LÃ¼th[1], J. Pozojevic[1], R. Tse[1], C. Reyes[1], H. Busch[2,3], A. Westenberger[1], C. Klein[1], J. Trinh[1]; [1]Inst. of Neurogenetics, LÃ¼beck, Germany, [2]LIED, LÃ¼beck, Germany, [3]Inst. for Cardiogenetics, LÃ¼beck, Germany
Disclosure Block: **J. LaÃ:** None.
Purpose: Establishment of long-read single-molecule Nanopore sequencing for the detection of i) repeat number, ii) methylation and iii) transcript expression of the SINE-VNTR-Alu (SVA) in the *TAF1* gene in X-linked dystonia-parkinsonism (XDP) patients. **Background:** XDP is an adult-onset neurodegenerative disorder characterized by rapidly progressive dystonia and parkinsonism. XDP is caused by a single founder mutation: an SVA retrotransposon insertion in the *TAF1* gene. Furthermore, the length of the polymorphic $(CCCTCT)_n$ domain within the *TAF1* SVA retrotransposon acts as a genetic modifier of disease onset and expressivity in XDP.
Methods: To first determine repeat number of the polymorphic $(CCCTCT)_n$ SVA domain, we used DNA derived from blood of patients with XDP (n=39). PCR amplicons of the *TAF1* SVA insertion were multiplexed with barcodes and sequenced using Nanoporeâ□□s LSK109. The hexanucleotide repeat number was obtained with NCRF. Fragment analysis was performed as a validation step for repeat detection. Secondly, to detect methylation, we used a Cas9-targeted enrichment approach with custom-designed gRNAs on DNA from blood, iPSC and brain of one patient and one control. CpG methylation detection spanning the *TAF1* SVA was performed with Nanopolish. Lastly, we used targeted IDT xGen Lockdown Probes SMARTer PCR cDNA Synthesis to obtain transcripts within the SVA on cDNA from iPSC and brain of one patient and one control. All libraries were sequenced on a R9 MinION flow cell.
Results: High concordance was observed for median repeat numbers of the $(CCCTCT)_n$ SVA domain using Nanopore sequencing and fragment analysis in blood-derived DNA (n=38 patients). The median CpG methylation frequency (MF) across blood, iPSC and brain tissue-derived DNA in the SVA is 0.912 (SD: 0.109). Likewise, the median MF up- and downstream of the SVA is 0.938 (SD: 0.124). CpG methylation in downstream enhancers flanking the SVA has a similar MF in patient with XDP compared to the control. 72 reads with the repeat polymorphism were detected in the cDNA transcripts. The average repeat number in cDNA transcripts is similar to the genomic DNA.
Conclusion: Repeat numbers and epigenetic modifications can be reliably detected using Nanopore sequencing. The insertion of the SVA retrotransposon in the *TAF1* gene does not lead to patient-specific differences of CpG MF and is linked to expression of novel transcripts containing an average of 44 hexanucleotide repeats.

PrgmNr 2574 - Exome sequencing in 200 intellectual disability/autistic patients: new candidates and atypical presentations

View session detail

Author Block: **L. Bruno**[1], F. Valentino[1], G. Doddato[1], A. Giliberti[1], R. Tita[2], S. Resciniti[1], C. Fallerini[3], m. bruttini[4], C. Lo Rizzo[2], M. Mencarelli[2], F. Mari[5], A. Pinto[2], F. Fava[5], M. Baldassarri[1], A. Fabbiani[5], V. Lamacchia[5], E. Benetti[6], K. Zguro[6], S. Furini[6], A. Renieri[5], F. Ariani[5]; [1]Med. Genetics, Univ. of Siena, Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies, Univ. of Siena, Siena, Italy, [2]Genetica Medica, Azienda Ospedaliera Univ.ria Senese, Siena, Italy, [3]Med. Genetics, Univ. of Siena, Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies, Univ. of Siena, SIENA, Italy, [4]Med. Genetics, Univ. of Siena, Genetica Medica, Azienda Ospedaliera Univ.ria Senese, Siena, Italy, [5]Med. Genetics, Univ. of Siena, Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies, Univ. of Siena, Genetica Medica, Azienda Ospedaliera Univ.ria Senese, Siena, Italy, [6]Med Biotech Hub and Competence Ctr., Dept. of Med. Biotechnologies, Univ. of Siena, Siena, Italy

Disclosure Block: **L. Bruno:** None.

Intellectual disability (ID) and autism spectrum disorder (ASD) belong to neurodevelopmental disorders and occur in ~1% of the general population. Due to disease heterogeneity identifying the etiology of ID and ASD remains challenging. Exome Sequencing (ES) offers the opportunity to rapidly identify variants associated with these two entities that often co-exist. Here, we performed ES in a cohort of 200 patients: 84 with isolated ID and 116 with ID and ASD. We identified 41 pathogenic variants with a detection rate of 22% (43/200): 39% in ID patients (33/84) and 9% in ID/ASD patients (10/116). Most of the causative genes are genes responsible for well-established genetic syndromes that have not been recognized for atypical phenotypic presentations. Six genes emerged as new candidates: BOD1, CACNA2D1, COBL, GPR14, KCNJ16, and NEURL4. In conclusion, this study reinforces the importance of ES in the diagnosis of ID/ASD and underlines that â□□reverse phenotypingâ□□ is fundamental to enlarge the phenotypic spectra associated with specific genes.

PrgmNr 2575 - Exploring the phenotypic spectrum of ARID1B-related disorders: Further two case reports

View session detail

Author Block: **S. Lemos Ferreira**[1], M. Santos Melo[1], J. Catanho da Silva[1], R. Pina[2], F. Tapadinhas[3], N. Silva[3], M. Amorim[1]; [1]ServiÃ§o de GenÃ©tica MÃ©dica, Hosp. Dona EstefÃ¢nia, Centro Hosp.ar e UniversitÃ¡rio de Lisboa Central, Lisboa, Portugal, [2]Unidade de Endocrinologia, ServiÃ§o de Pediatria MÃ©dica, Hosp. Dona EstefÃ¢nia, Centro Hosp.ar e UniversitÃ¡rio de Lisboa Central, Lisboa, Portugal, [3]ServiÃ§o de Pediatria, Hosp. Distrital de Faro, Faro, Portugal
Disclosure Block: **S. Lemos Ferreira:** None.
Introduction: Haploinsufficiency of *ARID1B* gene disrupts the BAF complex, affects neural development, and it is one of the main causes of intellectual disability. Clinical evidence suggests that the spectrum of ARID1B-Ârelated disorders (ARID1B-ÂRD) is highly variable and range from non-syndromic intellectual disability (ARID1BÂ-ID) to CoffinÂ-Siris syndrome (ARID1BÂ-CSS, OMIM #135900). Typically, patients with ARID1B-CSS display hypertrichosis, sparse scalp hair, thick eyebrows, long eyelashes, thick alae nasi, long and broad philtrum, and aplasia or hypoplasia of the distal phalanx or nail of the fifth finger/toe. Other features present in ARID1BÂ-RD include structural malformations (cardiac, central nervous system, musculoskeletal, gastrointestinal and genitourinary), laryngomalacia, feeding difficulties, growth problems, behavioural issues, and sensory impairment. Here, we present two cases that contribute to an increasing understanding of the clinical spectrum of ARID1B-ÂRD. **Case reports:** One patient, a 8-year-old male, presented with intellectual disability, attention deficit hyperactivity disorder, facial dysmorphisms (posterior low hairline, low-set and posterior rotated ears, long eyelashes, thick eyebrows, abnormal nasal bridge, short nose), short stature, delayed bone age, laryngomalacia, *pectus carinatum*, *pes planus*, strabismus, and nystagmus. Exome sequencing identified a likely pathogenic variant in *ARID1B* (NM_001374820.1):c.1293_1320del p.(Ala432Metfs*11). Parent testing is ongoing. A second patient, a 3-year-old male, was referred for development delay, hypotonia, minor dysmorphisms (long eyelashes and full lips), deafness, laryngomalacia, food aversion, pyloric stenosis, umbilical hernia, recurrent infections, autistic traits, sleep disturbance and brain anomalies (midbrain and corpus callosum dysmorphisms). Exome sequencing identified a *de novo* pathogenic variant in *ARID1B* (NM_020732.3):c.5890G>T p.(Glu1964*). **Discussion:** Both patients do not have a fully recognizable phenotype associated with ARID1B-ÂCSS. Also, the overall phenotypes reflect the wide variability of ARID1B-ÂRD, and are consistent with the reported cases. To date, there is no established genotypeÂ-phenotype correlation within the spectrum of ARID1BÂ-RD and the severity of the phenotype. **Conclusions:** With this work, we stress that ARID1BÂ-RD spectrum is extremely wide and ARID1B-CSS is only the tip of the iceberg, as highlighted by our patients. We also bring to attention the importance of reverse phenotype when evaluating a patient with a neurodevelopment disorder and a pathogenic variant in *ARID1B* gene.

PrgmNr 2576 - Factors influencing the chance of a genetic diagnosis in >13,000 individuals with developmental disorders

View session detail

Author Block: **P. Campbell**[1,2], V. K. Chundru[1], H. V. Firth[2], D. R. FitzPatrick[3], C. F. Wright[4], M. E. Hurles[1], H. C. Martin[1]; [1]Wellcome Sanger Inst., Cambridge, United Kingdom, [2]Cambridge Univ. Hosp. Fndn. Trust, Addenbrooke's Hosp., Cambridge, United Kingdom, [3]Univ. of Edinburgh, Edinburgh, United Kingdom, [4]Univ. of Exeter, Exeter, United Kingdom
Disclosure Block: **P. Campbell:** None.
The Deciphering Developmental Disorders (DDD) study is a UK-wide project that has, over the past 10 years, collected genomic and phenotypic data for >13,000 children with severe undiagnosed developmental disorders and their parents. Using whole exome sequencing and exon-arrayCGH, the overall diagnostic yield to date is around 31% - similar to that reported for comparable cohorts. Some factors are known to influence the chance of an individual patient receiving a diagnosis, such as whether they are sequenced in a trio. However, other factors influencing the chance of reaching a genetic diagnosis have not yet been explored in depth. We performed multiple logistic regression to see how various demographic, clinical, phenotypic, maternal and ancestral factors affect the chances of gaining a diagnosis within the DDD study. This model explained ~13% of the variance in chance of getting a diagnosis. We find that trio status has the largest impact on the chance of gaining a diagnosis (OR: 4.5, 95% CI 4.0 - 5.1). Other factors significantly increasing the chance of diagnosis include: being the only affected family member (OR: 1.7, 95% CI: 1.6 - 1.9), having severe developmental delay (OR: 2.4, 95% CI: 2.1 - 2.7) and having increasing number of organ systems affected (OR: 1.1, 95% CI: 1.1 - 1.1). In the DDD cohort, male sex (OR: 0.7, 95% CI: 0.7 - 0.8), increasing consanguinity (OR: 0.7, 95% CI: 0.6 - 0.8) and African ancestry (OR: 0.5, 95% CI: 0.3 - 0.8) reduce the odds of finding a diagnosis. We find evidence of a non-linear interaction between trio status and African ancestry (P

PrgmNr 2577 - Human patient SFPQ homozygous mutation is found deleterious for brain and motor development in a zebrafish model

View session detail

Author Block: **S. Efthymiou**[1], P. M. Gordon[2], V. Salpietro[1], T. Fielding[2], E. Borgione[3], C. Scuderi[3], H. Houlden[1], C. Houart[2]; [1]UCL Inst. of Neurology, London, United Kingdom, [2]Ctr. for Dev.al Neurobiology and MRC Ctr. for Neurodevelopmental Disorders, IoPPN, Guy's Campus, King's Coll. London, London, United Kingdom, [3]Oasi Res. Inst.-IRCCS, Troina, Italy
Disclosure Block: **S. Efthymiou:** None.
SFPQ (Splicing factor proline- and glutamine-rich) encodes a DNA and RNA binding protein involved in transcription and pre-mRNA splicing. Loss of *sfpq* causes defects in neuronal development, and rare variants in this gene have been previously implicated in neurodegenerative disorders. In this study, we identified a homozygous p.Ser660Asn variant in *SFPQ* through whole exome sequencing (WES) of an Italian family affected with a complex neurological phenotype characterized by intellectual disability, peripheral neuropathy, vacuolar myopathy, parkinsonian features and neuroradiological anomalies resembling neurodegeneration with brain iron accumulation (NBIA). Rescuing a zebrafish *SFPQ* homozygous null mutant with this human SFPQ^{S660N} variant revealed robust defects in the developing central nervous system (CNS) of the embryos, including microcephaly, misfolding of the posterior brain neuroepithelium and abnormal branching of the motor axons innervating body muscles. The defects identified in the SFPQ^{S660N} zebrafish model provide evidence supporting a contribution of SFPQ^{S660N} to the patientsâ complex pathology and suggest that *SFPQ* mutations may contribute to different broad clinical and neuroradiological degenerative features in humans.

PrgmNr 2578 - Identification of rare copy number variants in a family with high prevalence of psychosis

View session detail

Author Block: **F. Francis**; UCR, San Jose, Costa Rica
Disclosure Block: **F. Francis:** None.
Psychotic disorders are a relevant subject of public health because of their prevalence, economic significance and emotional suffering to families and patients. However, due to their complex disease nature, among others, they are difficult to study. Currently, Copy Number Variants (CNVs), have become more relevant, not only for their contribution to the genome variability among individuals, but also because some of these variants increase the risk of SZ, BPD and autism spectrum disorders (ASD). These variants tend to have much larger effect size than SNPs, however the bioinformatic complexity for the analysis of these variants remains high. In addition, rare variants, are most commonly found in families or individual cases, usually have a larger effect, and are of recent origin, which gives them an important role to characterize the disease. Despite this, those variants are underrepresented in the current databases. We identified a family in Costa Rica with 219 individuals (121 women, 96 men), of whom 56 have presented some psychotic disorder (31 women, 25 men) such as bipolar disorder, unspecified affective psychosis, schizo-affective disorder and unspecified psychosis according to the DSM-IV, confirmed by a personal interview and the process of best estimate diagnosis. Since studies with related individuals help reduce the sample size required to maintain statistical power, increasing the probability of capturing the association between rare variants and phenotype, we analyzed 114 whole genome sequences of individuals from the family described above. The sample includes 48 cases (27 women and 21 men) and 66 controls (38 women and 28 men), the sequence is paired ends, with a minimum read length of 2x100bp and minimum of 30x average coverage. We present the workflow used to identify a wide range of CNVs using a combination of software and algorithms to estimate CNVs breakpoints. The competencies of 15 CNV detection software packages, that are currently in use, were analyzed (CLAMMS, ReadDepth, CNAseg, XHMM, cn.mops, Seq Gene, SeqCNV, BreakDancer, Hadoop CNV, SurVIndel, GATK, Genome STRiP, RDXplorer, LUMPY, CNVnator), of these, CNVnator and GATK were selected, for various reasons. Currently the pipeline has a total time of 80m8.083s, paralleling by sample and not by Jobs, with GNU parallel and Samtools, instead of Sambamba. The pipeline was standardized with 6 .cram files, in which the average RD per Bin was 174.4 Â± 17.0 before the GC correction and we estimated in average 1991 Â± 127 CNVs per sample. This is the first study to identify CNV of risk for psychotic disorders in a Costa Rican population.

PrgmNr 2579 - Infantile ascending spastic paraplegia linked to *ALS2*pathogenic variant

View session detail

Author Block: **L. Schottlaender**[1,2,3], H. Amartino[4], S. Efthymiou[5], M. Donato[4], E. Cali[6], V. Chelban[7,6], F. Heredia[8], A. Berardo[9], S. JuÃ¡rez PeÃ±alva[1,10,11], L. LeÃ³n Cejas[12], A. Pardal[12], J. Vandrovcova[13], H. Houlden[6], C. Marchesoni[12], R. Reisin[12], IMeG-study group; [1]Inst. de medicina genÃ³mica (IMeG), Hosp. Univ.rio Austral, Univ. Austral, Pilar, Argentina, [2]Inst. de Investigaciones en Med. Traslacional (IIMT), CONICET-Univ. Austral, Pilar, Argentina, [3]Servicio de NeurologÃa, Hosp. Univ.rio Austral, Univ. Austral, Pilar, Argentina, [4]Servicio de NeurologÃa Infantil, Hosp. Univ.rio Austral, Univ. Austral, Pilar, Argentina, [5]UCL Inst. of Neurology, London, United Kingdom, [6]UCL Queen Square Inst. of Neurology, London, United Kingdom, [7]Chisinau, [8]Dept. of Neurology, Hosp. BritÃ¡nico,, CABA, Argentina, [9]Sanatorio Allende Nueva CÃ³rdoba, CÃ³rdoba, Argentina, [10]Servicio de GenÃ©tica, Hosp. Univ.rio Austral, Univ. Austral, Pilar, Argentina, [11]Div. of Fetal Med., Hosp. Univ.rio Austral, Pilar, Argentina, [12]Dept. of Neurology, Hosp. BritÃ¡nico, CABA, Argentina, [13]London's Global Univ., United Kingdom
Disclosure Block: **L. Schottlaender:** None.

ALS2-related disorders are rare monogenic disorders that have variable affection of upper and lower motor neurons usually presenting with upper motor neuron and cranial nerve involvement. Most patients are described within the clinical spectrum of motor neuron disorders subdivided into three main diseases: Infantile ascending hereditary spastic paraplegia (IAHSP), juvenile primary lateral sclerosis and juvenile amyotrophic lateral sclerosis. However, other forms such as dystonia have been described. *ALS2*-related disorders are caused bi-allelic variants in the gene *ALS2* encoding for alsin mostly described in consanguineous communities.

Here, we describe a family with two affected members presenting with IAHSP from the indigenous toba community from the Chaco region of Argentina. We present two siblings that began early in life with delayed motor milestones. One male and one female, both had late sitting and walking, and lost this ability after a few years. They later evolved to tetraparesis and bulbar involvement with anarthria. Although given the language difficulties the cognitive assessment is limited, they understand complex commands, they know how to read and manage independently with telephones and computers. The older sibling, currently 21 years old has mild swallowing difficulties, but his younger 14-year-old sister does not have swallowing difficulties. They belong to the indigenous toba community, parents are consanguineous and have other unaffected siblings.

Through a collaborative next generation sequencing study a homozygous variant in *ALS2* was detected in both patients: NM_020919:c.T2531A(p.L844H). This variant is predicted to be pathogenic and previously reported in another patient from Argentina. They are not known to be from the same community but both parents are originally from a small remote town. This second family had two first cousins affected, and we are currently following up the family origin to determine if they belong to a related indigenous community.

In summary, we describe an interesting consanguineous toba family with 2 members affected by a homozygous pathogenic variant in *ALS2* presenting in childhood with IAHSP. Diagnosis for this family was established after two decades of symptom onset and was essential for genetic counselling. Achieving a molecular diagnosis in such rare disorders is essential for family planning and the development of biomarkers and targeted therapies.

PrgmNr 2580 - Inherited variants in *CHD3* demonstrate variable expressivity in Snijders Blok-Campeau syndrome

View session detail

Author Block: **J. den Hoed**[1], J. van der Spek[2], L. Snijders Blok[3], A. J. M. Dingemans[2], D. Schijven[1], T. S. Barakat[4], E. M. Bebin[5], S. Beck-WÃ¶dl[6], G. Beunders[7], N. J. Brown[8], T. Brunet[9], H. G. Brunner[2], P. Campeau[10], G. Cuturilo[11], C. Gilissen[2], T. B. Haack[6], R. A. Husain[12], B. Kamien[13], S. Chern Lim[14], L. Lovrecic[15], J. Magg[6], A. Maver[15], D. C. Monteil[16], C. W. Ockeloen[2], C. Richmond[14], A. Riess[6], E. M. C. Schwaibold[17], M. E. H. Simon[18], S. Spranger[19], T. Tan[14], M. L. Thompson[20], B. de Vries[21], E. J. Wilkins[14], M. H. Willemsen[2], L. E. Vissers[22], S. E. Fisher[23], T. Kleefstra[24]; [1]Max Planck Inst. for Psycholinguistics, Nijmegen, Netherlands, [2]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [3]Radboud Univ Med. Ctr., Nijmegen, Netherlands, [4]Erasmus MC Univ. Med. Ctr., Rotterdam, Netherlands, [5]Univ. of Alabama, Birmingham, AL, [6]Inst. of Med. Genetics and Applied Genomics, Univ. of TÃ¼bingen, TÃ¼bingen, Germany, [7]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [8]Univ. of Melbourne, Melbourne, Australia, [9]Technical Univ. Munich, Munich, Germany, [10]Montreal, QC, Canada, [11]Univ. Children's Hosp. Belgrade, Belgrad, Serbia, [12]Jena Univ. Hosp., Jena, Germany, [13]King Edward Mem. Hosp., Subiaco, Australia, [14]Victorian Clinical Genetics Services, Victoria, Australia, [15]UMC Ljubljana, Ljubljana, Slovenia, [16]Naval Med. Ctr. Portsmouth, Portsmouth, VA, [17]Heidelberg Univ., Heidelberg, Germany, [18]Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, [19]Praxis fÃ¼r Humangenetik-Bremen, Bremen, Germany, [20]Hudson Alpha Inst. for Biotechnology, Huntsville, AL, [21]Nijmegen, Netherlands, [22]Radboudumc, Nijmegen, Netherlands, [23]Max Planck Inst for Psycholinguistics, Nijmegen, Netherlands, [24]Radboud umc, Nijmegen, Gelderland, Netherlands

Disclosure Block: **J. den Hoed:** None.

Interpretation of next-generation sequencing data of individuals with an apparent sporadic neurodevelopmental disorder (NDD) often focusses on *de novo* variants in genes known for autosomal or X-linked dominant inheritance, or on bi-allelic variants in genes described with recessive inheritance. As a consequence, inherited variants in genes associated with dominant disorders may be overlooked when the transmitting parent is clinically unaffected. While *de novo* variants explain a substantial proportion of cases with NDDs, still a significant number remains undiagnosed. Various factors may explain the inability to diagnose these individuals, including coding variants associated with reduced penetrance and variable expressivity. Here, we characterized 18 families with inherited missense or protein-truncating variants (PTV) in *CHD3*, a gene in which *de novo* variants were recently described to cause a neurodevelopmental syndrome, characterized by intellectual disability, speech delay and recognizable facial features (Snijders Blok-Campeau syndrome; MIM#618205). With in-depth analysis, including computational facial and human phenotype ontology-based comparisons, we demonstrated that the phenotypic features of probands with inherited *CHD3* variants overlap with the phenotype previously associated with *de novo* variants in the gene, while carrier parents are mildly or not affected, suggesting variable expressivity. Additionally, we demonstrated, both in healthy carriers and in a related affected proband with a *CHD3* PTV, similarly reduced expression levels of CHD3 protein in cells, suggesting that allelic compensation is unlikely to be an underlying mechanism. Our results point to a role of inherited variation in well-described NDDs previously associated almost exclusively to *de novo* variants, a finding that is critical for genetic counseling and warrants further investigation towards understanding the broader contributions of such variation to the landscape of human disease.

PrgmNr 2581 - Methylation status of chromosome 5 in *Cri du Chat*Syndrome: new insights for clinical variability?

View session detail

Author Block: **V. Almeida**[1], G. Carvalho[1], S. Chehimi[1], Y. Gasparini[1], A. Nascimento[1], C. Kim[2], L. Kulikowski[1]; [1]Faculdade de Med. da Univ.e de SÃ£o Paulo, Sao Paulo, Brazil, [2]Inst. da Crianca, Hosp. das Clinicas, Faculdade de Med., Sao Paulo, Brazil
Disclosure Block: **V. Almeida:** None.
Cri Du Chat Syndrome (SCDC) or 5p- Syndrome (OMIM #123450) is characterized by a genomic loss in the short arm of chromosome 5 and variable clinical manifestations, that include highâ□□pitched cry in newborns, facial asymmetry, microcephaly, hypotonia, ocular hypertolorism, micrognathia, neurological and behavioral changes. Different cytogenomics rearrangements, family history and environmental factors may hinder genotype - phenotype association. Thus, phenotypic variability in this syndrome may not be limited only to variations in gene structure, such as deletions, duplications, inversions, insertions and translocations, being possible that others mechanisms related to promoters activation or inactivation and/or exons of actively transcribed genes, including DNA methylation, which occurs mainly in the â□□CpG Islandsâ□□. In this sense, we studied the profile of methylation status of chromosome 5 in 10 patients and investigated regions around the chromosomal breakpoint at 5p to verify the differentially methylated status. DNA samples from seven SCDC patients with similar breakpoints and three previously genotyped control samples (HumanCyto850K BeadChip) were evaluated using the chip Infinium MethylationEPIC BeadChip (Illumina, Inc., San Diego, CA). The analysis was performed with RStudio software. The patient's heatmaps show greater heterogeneity of methylated probes for the 5p region compared to control samples also the results reveals that some genes important to the SCDC phenotype have hypermethylated methylation status, consistent with the patient's phenotype.

PrgmNr 2582 - Modeling a neuropathy-associated *GARS1* mutation in *C. elegans*

View session detail

Author Block: **J. Pierluissi**[1], A. Antonellis[2,3]; [1]Dept. of Human Genetics, Univ. of Michigan, Ann Arbor, MI, [2]Univ. of Michigan, Ann Arbor, MI, [3]Dept. of Neurology, Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **J. Pierluissi:** None.
Charcot-Marie-Tooth disease (CMT) is a heritable peripheral neuropathy that is characterized by motor and sensory defects in the distal extremities. Mutations in five genes encoding aminoacyl-tRNA synthetases (ARSs)â□□a family of enzymes that ligate amino acids to cognate tRNA molecules during the beginning stages of protein translationâ□□have been implicated in dominantly inherited CMT. One of these implicated enzymes is glycyl-tRNA synthetase (GARS1), the enzyme responsible for ligating glycine to cognate tRNA molecules. It has been demonstrated that there is wide allelic and clinical heterogeneity of *GARS1*-mediated neuropathy. We recently reported a 12-base-pair, in-frame deletion in *GARS1*, (E245_Q248; or Î□ETAQ) in a patient with a severe form of peripheral neuropathy that is similar to infantile-onset spinal muscular atrophy. The Î□ETAQ mutation ablates enzyme activity *in vitro*, reduces viability in yeast complementation assays, and is dominantly toxic to mouse neurons. To further determine the pathological significance of the Î□ETAQ mutation and to establish a pipeline by which disease-causing ARS mutations may be systematically studied in a robust model organism, we employed a CRISPR/Cas9 method to generate a *C. elegans* model of *GARS1*-mediated disease. Here, we provide characterization of the first *C. elegans GARS1*-mediated neuropathy model, which supports a loss-of-function effect of the patient variant. Pharmacological characterization of the neuromuscular junction of mutant worms indicates a degenerative defect in synaptic transmission, specifically implicating the presynaptic terminal of the cholinergic motor neurons. Here, I will present our unpublished data and plans to improve the phenotype observed in worm toward developing patient therapeutics. This work contributes to our understanding of the role of *GARS1* in peripheral neuropathy and provides a framework for studying the pathogenicity of other ARS mutations of interest.

PrgmNr 2583 - Monoallelic variants in *TFAP2E* are implicated in craniofacial and central nervous system anomalies

View session detail

Author Block: **J. Kalanithy**[1,2], G. C. Dworschak[1,2,3], J. A. Rosenfeld[4,5], E. Mingardo[1,2], J. D. Stegmann[1,2], A. C. Hilger[1,2,3], W-H. Tan[6], S. A. Coury[6], A. Woerner[7], J. Sebastian[7], P. A. Levy[8], T. T. Lindenberg[2], Ã. Yilmaz[2], A. Becker[9], H. Thiele[10], J. E. Posey[11], J. R. Lupski[12], H. Reutter[1,13], B. Odermatt[2], W. M. Merz[14]; [1]Inst. of Human Genetics, Med. Faculty, Univ. of Bonn, Bonn, Germany, [2]Inst. of Anatomy and Cell Biology, Med. Faculty, Univ. of Bonn, Bonn, Germany, [3]Dept. of Pediatrics, Childrenâ□□s Hosp. Med. Ctr., Univ. Hosp. Bonn, Bonn, Germany, [4]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [5]Baylor Genetics Lab., Houston, TX, [6]Div. of Genetics and Genomics, Boston Childrenâ□□s Hosp, Boston, MA, [7]Div. of Genetic and Genomic Med., Dept. of Pediatrics, UPMC Children's Hosp. of Pittsburgh, Pittsburgh, PA, [8]Dept. of Pediatrics, Children's Hosp. at Montefiore, Bronx, NY, [9]Section for Translational Epilepsy Res., Dept. of Neuropathology, Bonn, Germany, [10]Univ. of Cologne, Cologne, NRW, Germany, [11]Baylor Coll. of Med., Houston, TX, [12]Baylor Col Med, Houston, TX, [13]Dept. of Neonatology and Pediatric Intensive Care, Childrenâ□□s Hosp. Med. Ctr., Univ. Hosp. Bonn, Bonn, Germany, [14]Dept. of Obstetrics and Prenatal Med., Univ. Hosp. Bonn, Bonn, Germany
Disclosure Block: **J. Kalanithy:** None.

TFAP2E encodes the transcription factor AP-2. The potential role of TFAP2E (AP-2) transcription factor in development and the formation of the central nervous system (CNS) and craniofacial structures has not been systematically investigated . Here, we report 3 individuals with novel heterozygous missense variants in *TFAP2E*, all presenting with CNS malformations and craniofacial anomalies. For one family, parental samples were available to confirm *de novo* occurrence of the variant allele. Variants were identified by exome sequencing (ES). We initiated functional characterization in zebrafish larvae using splice and translational blocking morpholinos (MO) in different transgenic zebrafish lines. Fluorescent signals of neuronal, glial and neural crest cells at 2 days post fertilization (dpf) were analysed *in vivo* using a twoâ□□photon point scanning microscope and subsequent 3D modelling with IMARIS analysis software. Alcian blue cartilage staining was performed at 3 dpf. Shared clinically observable phenotypic features among patients comprised malformations of the CNS (3/3) and craniofacial anomalies (3/3) including micrognathia or microcephaly. MO knockdown of *tfap2e* in zebrafish larvae showed microcephaly, hydrocephalus und craniofacial anomalies and thus partially recapitulate the human phenotype. Major effects on vasculogenesis were excluded after MO injections into vascular reporter lines, suggesting the tissue specificity of the observed effects. Our findings implicate *TFAP2E* in CNS and neural crest development and suggest *TFAP2E* as a novel candidate gene for human CNS malformations and craniofacial anomalies. Grant: J.C.K: BonnNI grant Q614.0754 and BONFOR grant O-167.0023. HR and HT: German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) (RE 1723/5-1, and TH 1327/2-1). J.R.L: Jointly funded National Human Genome Research Institute (NHGRI) and National Heart, Lung, and Blood Institute (NHLBI) grant to the Baylor-Hopkins Center for Mendelian Genomics [UM1 HG006542].

PrgmNr 2584 - NeuroWES-Macedonia: genetic diagnoses for complex syndromic cases with neurodevelopmental disorders

View session detail

Author Block: **S. Trajkova**[1], E. Sukarova-Angelovska[2], A. Petlichkovski[3], B. Gagachovska[4], E. Spritova[5], V. Trajkova[6], G. Kiteva-Trencevska[7], S. Cardaropoli[1], L. Pavinato[8], V. Pullano[1], C. Giovenino[9], S. Carestiato[10], T. Pippucci[11], P. Dimartino[12], S. De Rubeis[13], J. D. Buxbaum[13], A. Brusco[14]; [1]Dept. of Med. Sci., Univ. of Turin, Turin, Italy, [2]Dept. of Endocrinology and Genetics, Univ. Clinic for Children's Diseases, Faculty of Med., Univ. 'Sv. Kiril I Metodij', Skopje, Macedonia, The Former Yugoslav Republic of, [3]Inst. for Immunobiology and Human Genetics, Faculty of Med., Univ. 'Sv. Kiril I Metodij', Skopje, Macedonia, The Former Yugoslav Republic of, [4]Clinic for Psychiatric Disease and Mental Hlth., Clinical Hosp., Bitola, Macedonia, The Former Yugoslav Republic of, [5]Univ. Clinic for Children's Diseases, Faculty of Med., Univ. 'Goce Delcev', Shtip, Macedonia, The Former Yugoslav Republic of, [6]Univ. Clinic of Neurology, Faculty of Med., Univ. 'Goce Delcev', Shtip, Macedonia, The Former Yugoslav Republic of, [7]Univ. Clinic of Neurology, Faculty of Med., Univ. 'Sv. Kiril I Metodij', Skopje, Macedonia, The Former Yugoslav Republic of, [8]Univ. of Turin, Turin, Italy, [9]UniversitÃ degli Studi di Torino, Torino, [10]Univ. of Turin, Torino, Italy, [11]Sant'Orsola Malpighi Univ. Hosp., Bologna, BO, Italy, [12]Dept. of Med. and Surgical Sci., Univ. of Bologna, Bologna, Italy, Bologna, Italy, [13]Icahn Sch. of Med. at Mount Sinai, New York, NY, [14]CittÃ della Salute e della Scienza Univ. Hosp., Torino, Italy, Turin, Italy

Disclosure Block: **S. Trajkova:** None.

Neurodevelopmental disorders (NDD) is a broad term that includes Autism Spectrum Disorders (ASD), intellectual disability (ID), and epilepsy. Advances in sequencing technology have greatly expanded our understanding of the NDD genetics. More than 100 rare Mendelian diseases have been associated with NDDs. Given the high-level of phenotypic and genotypic heterogeneity, the correct molecular diagnosis requires not only up-to-date sequencing/bioinformatic tools but also a detailed clinical investigation. We collected and deeply phenotyped 212 pediatric NDD cases from Macedonia (NeuroWES-Macedonia) affected by ASD/ID (82/212, 39%), complex syndromes (74/212; 35%), and isolated ASD (55/212; 26%). We performed trio-Whole Exome Sequencing (WES) within the Autism Sequencing Consortium (ASC) international collaborative network. Currently, we analysed 151 trios. We found pathogenic/likely pathogenic variants in 39 (26%), with the highest diagnostic rate among syndromic cases (21/44; 47%), and the lowest in the isolated ASDs (3/31; 9%). Most causative variants were *de novo* (dn, 25/39, 65%); the remaining autosomal recessive (7/39, 18%) or X-linked inherited (7/39, 18%). Among syndromic cases we found pathogenic variants in exceedingly rare disease-associated genes: truncating variants in *HIST1H1E* (MIM 617537)*, PQBP1 (*MIM 309500), *USP9X* (MIM 300968) and missense variants in *EZH2 (*MIM 277590), *HDAC8* (MIM 300882*)*, *PHF6 (*MIM 301900). Being these genes involved in chromatin-remodeling pathways, an â□□epigenetic signatureâ□□ analysis is ongoing within the EpiSign consortium. Dual molecular diagnoses were found in 2 cases (4%). In one family, we found two similarly affected brothers: one had a causative *dn* frameshift in *TRIP12* (*604506)*,* the other the dn recurrent missense p.A1728V in *FBN1*. In a case with microcephaly, short stature, and impaired glucose metabolism (MIM 616817), we found a homozygous p.R658H in *PPP1R15B,* the third missense pathogenic variant described in this gene. We also found 4 novel strong candidate NDD genes: a *dn* frameshift in *PRMT8* (*610086); a *dn* missense in *PHLPP1* (*609396); a *dn* missense in *OSBPL8* (* 606736) in a proband with a Lenz-Majewski-like clinical presentation; finally, a homozygous missense in *AARSD1* likely disrupting the function of this alanyl-tRNA editing activity protein highly related to brain myelinations processes. We realized that deep phenotyping combined with trio-WES genetic data is a key strategy to solve undiagnosed NDD cases, and go beyond clinical complexity. We suggest novel possible NDD genes which we will study in the near future.

PrgmNr 2585 - Paediatric neurogenetic disorders in Bangladesh: First tier exome sequencing in previously unexplored population

View session detail

Author Block: **H. Morsy**[1], S. Efthymiou[1], M. Zaman[2], V. Salpietro[1], SYNaPS Study Group, J. Vandrovcova[1], S. Banu[2], H. Houlden[1]; [1]UCL Inst. of Neurology, London, United Kingdom, [2]Khan Shishu Hosp. and ICH, Dhaka, Bangladesh
Disclosure Block: **H. Morsy:** None.
Background: Peadiatric neurogenetic disorders are highly heterogenous both on phenotypic and molecular basis. Identification of underlying etiology is highly significant for clinical care of the patients. In Bangladesh, these disorders have not been previously characterized. We describe clinical and genomic data from a large cohort of Bangladesh comprising 230 families recruited from 2018 to 2020.
Methods: All patients were deeply phenotyped and assigned a diagnostic group based on Genomics England clinical classification. Whole Exome Sequencing (WES) was applied as a first-tier diagnostic testing for 230 affected individuals, of which 78 probands were sequenced as duo/trio. Exome data were screened for ultra-rare deleterious variants in known and candidate genes in Neurology and Neurodevelopmental Panel according to Genomics England. Applying additional bioinformatics pipelines such as SpliceAI and Conifer for calling of splice altering variants, CNVs and small Indels is currently ongoing.
Results: Our cohort compromised patients mainly with Inherited epilepsy syndromes or Neurodevelopmental disorders with frequency of 47.8 % and 34.7%, respectively. More than 1/3 of the patients were younger than 2 years old and almost half of the cohort were born to first degree cousin parents. We had a definitive or likely genetic diagnosis at 31.3 % (72/230) of cases. Interestingly, more than 2/3 of the solved cases was caused by monoallelic variants (68.05%). Even more, ultra-rare monoallelic variants in our cohort which is highly consanguineous (49.1 %) compromised 52.46 % of all protein coding variants in Neurology and neurodevelopmental gene panel.
Conclusion: We describe the challenges facing WES data analysis in previously unexplored population and how good phenotyping and additional Bioinformatics pipelines can improve the diagnostic yield. We highlight the significant proportion of ultra-rare monoallelic deleterious variants in Paediatric neurogenetics disorders even in highly consanguineous population.

PrgmNr 2586 - Phenotypic characteristics of a family with distal arthrogryposis type 2B (Sheldon-Hall syndrome) due to*TNNT3*pathogenic variant

View session detail

Author Block: **D. GERMAIN**[1], I. DABAJ[2], R. Y. CARLIER[3], O. AILIOAIE[3], S. QUIJANO-ROY[3]; [1]Paris Saclay Univ., Montigny, France, [2]CHU Rouen, Rouen, France, [3]APHP - Paris Saclay Univ., Garches, France
Disclosure Block: **D. Germain:** None.
Background: Distal arthrogryposis (DA) is a group of rare, clinically and genetically heterogeneous disorders characterized by congenital contractures of the limb joints. Sheldon-Hall syndrome (SHS or DA2B) is a rare multiple congenital contracture syndrome characterized by contractures of the distal joints. SHS is intermediate to DA1 and DA2A (Freeman-Sheldon syndrome) and was found to be due to mutations in genes encoding the fast-twitch skeletal muscle contractile myofibers complex.
Patients and methods: A 16-year old boy with a history of congenital distal arthrogryposis presented to our multidisciplinary consultation with severe kyphoscoliosis and respiratory insufficiency. Both his mother and younger sister, who had not previously diagnosed, were found to have a milder phenotype. Diagnostic work up included physical exam, whole body muscular MRI (WBMRI) in all three patients, together with ENMG studies and muscle biopsy in the index case. DNA sequencing was used to confirm the clinical diagnosis. **Results:** Physical examination suggested the clinical diagnosis of SHS. No major abnormalities were found in WBMRI for all three individuals. Neurogenic changes were observed on needle EMG and on muscle biopsy in the index patient. DNA sequencing revealed a missense pathogenic variant in *TNNT3* (c.187C>T; p.Arg63Cys). This variant was found to segregate with the SHS phenotype in affected individuals. **Discussion:** This is the first report of neurogenic involvement in a patient with DA2B. A missense mutation was identified at codon 63 a mutational hotspot of the *TNNT3* gene. While *TNNT3* encodes the fast-twitch skeletal muscle contractile myofibers complex, our data suggests additional chronic nerve injury in this arthrogryposis syndrome.

PrgmNr 2587 - Recessive burden analysis and gene discovery in ~40,000 developmental disorder trios with diverse ancestries

View session detail

Author Block: **V. Chundru**[1], Z. Zhang[2], K. E. Samocha[1], R. Torene[2], M. E. Hurles[1], K. Retterer[2], H. C. Martin[1]; [1]Wellcome Sanger Inst., Hinxton, United Kingdom, [2]GeneDx, Gaithersburg, MD
Disclosure Block: **V. Chundru:** None.
Previous work in the Deciphering Developmental Disorders (DDD) study has estimated that ~3% of White British cases and ~30% of British South Asian cases could be attributed to recessive coding variants. The difference in burden of recessive coding variants between these populations is driven by differing rates of autozygosity due to consanguinity (Martin *et al*., Science, 2018). About half of this burden was previously attributed to as-yet-undiscovered genes, as gene discovery was limited due to lack of power in the predominantly European ancestry, and non-consanguineous DDD cohort.
Here, we present the largest multi-ancestry study to date for the discovery of novel recessive genes in developmental disorders (DDs). Combining the DDD study and GeneDx, we will analyse ~40,000 DD trios. However, rules around patient confidentiality preclude sharing complete raw data for patients. Despite this, we developed a federated approach to quality control, followed by data harmonisation based on comparing summary statistics, in order to be able to undertake mega-analyses of rare variant data from the two largest cohorts of DD patients.
Since properly calibrated burden analysis for recessive variants relies on population-specific allele frequency estimates, we use PCA and UMAP to examine fine-scale population structure. We identified 20 distinct ancestry groups, each with at least 100 trios, from European, African, South Asian, East Asian, South American, and Middle Eastern populations, which we are analysing individually. For example, we identify a homogeneous European group with ~25,000 trios (~0.08% with >1.56% autozygosity - corresponding to offspring of second cousins), a South Asian group with ~1,800 trios, and a Middle Eastern group with ~1,000 trios (both of which have >50% with >1.56% autozygosity). Using a burden testing framework accounting for levels of autozygosity and parental haplotype frequencies, we will estimate the excess of protein-truncating and damaging missense variants in each gene within each ancestry group. Summing these values across genes will allow us to estimate the recessive burden of DDs. We will leverage this larger sample size to discover new recessive genes for DDs through statistical burden analysis. Additionally, we will develop a simulation-based statistical enrichment test, weighting biallelic genotypes by their degree of exome-wide enrichment, to further increase our power to detect recessive DD genes.

PrgmNr 2588 - Recessive inactivating variants in *DMAP1* in an intellectual disability syndrome

View session detail

Author Block: **D. Li**[1], A. Sobering[2], Q. Wang[1], T. Duelund HjortshÃ¸j[3], S. Redler[4], K. Platzer[5], N. Giesbertz[6], R. Sulaiman[7], M. March[8], C. Applegate[9], A. Hamosh[10], E. J. Bhoj[11], F. S. Alkuraya[12], Y. Song[1], H. Hakonarson[8]; [1]The Children's Hosp. of Philadelphia, Philadelphia, PA, [2]St. George's Univ., Great River, NY, [3]Univ. Hosp. of Copenhagen, Copenhagen, Denmark, [4]Univ. of Leipzig Med. Ctr., Leipzig, Germany, [5]Leipzig, Germany, [6]Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, [7]King Faisal Specialist Hosp. and Res. Ctr., Riyadh, Saudi Arabia, [8]Children S Hosp. of Philadelphia, Philadelphia, PA, [9]Johns Hopkins Univ., Baltimore, MD, [10]Johns Hopkins Univ Sch Med, Baltimore, MD, [11]Philadelphia, PA, [12]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia

Disclosure Block: **D. Li:** None.

DMAP1 encodes a versatile protein that is involved in different complexes responsible for maintenance of DNA methylation when directly interacting with DNMT1, regulation of histone acetylation when coupled with the TRRAP/Tip60 complex, and catalysis of histone H2A.Z deposition in conjunction with the SRCAP complex. Despite DMAP1â□□s essential role in multiple transcriptional processes, it has not been implicated in human diseases. By exome sequencing and international matchmaking, we identified six unrelated individuals with a syndromic neurodevelopmental disorder carrying homozygous or compound heterozygous variants in *DMAP1*. Besides two truncating and one splice-altering variants, we uncovered five missense variants [c.291G>C, p.(Trp97Cys); c.338A>G, p.(His113Arg); c.371A>G, p.(Tyr124Cys); c.511T>C, p.(Phe171Leu), and c.581G>A, p.(Arg194Gln)] residing in or around the SANT domain, suggesting they may affect its interaction with DNA and/or histone. All six individuals have global developmental delay, intellectual disability, feeding difficulty, hypotonia, and craniofacial dysmorphic features, although the reported findings varied. Detailed clinical assessment demonstrated that an additional feature, such as microcephaly, short stature, seizures, and brain malformation, was concomitant in at least two individuals. Using the Gal4-UAS system in *Drosophila*, we expressed two independent *Dmap1* RNAis (RNAiBL63666, and RNAiv103734) under the control of *elav-Gal4*, a pan-neural driver, and found that both led to pupal lethality. Dissection of larval brains further demonstrated structural defects of the mushroom body (MB), the learning and memory center of fly. Compared to control, the volume of vertical lobe was significantly reduced after *Dmap1* knockdown, while the horizontal lobe was marginally affected, highlighting an underappreciated role of *Dmap1* in the MB development. Preliminary transcriptome analyses from *Drosophila* RNAi brains (N = 6) identified dysregulation of hundreds of genes implicated in transcription processing, neuronal function, and brain development. Together, our findings suggest that the *DMAP1* variants we discovered may exert a loss of function effect. To test our hypothesis, we are currently conducting rescue studies with the identified missense variants. In summary, we demonstrate that recessive variants in *DMAP1* are associated with a novel neurodevelopmental disorder.

PrgmNr 2589 - The phenotypic spectrum of *BRAT1*-related disorders : 51 additional cases and literature review

View session detail

Author Block: **C. ENGEL**[1], S. Valence[2], G. P. Delplancq[1], R. Maroofian[3], D. J. Morris-Rosendahl[4], A. A. Putoux[5], L. Pasquier[6], E. Boucher-Brischoux[1], C. Mignot[7], M. Doco-Fenzy[8], M. Rio[9], S. Strivastava[10], V. Salpietro[11,12], E. Valente[13], L. S. Penney[14], A. Brusco[15], M. GÃ©rard[16], J-H. Chae[17], S. Odent[18], M. Di Giacomo[19], R. Kaiyrzhanov[3], D. Kraus[20], I. Scheffer[21], J. G. Gleeson[22], M. S. Zaki[23], V. Baglioni[24], R. Ibadova[25], L. Van Maldergem[1], L. Burglen[26], J. Piard[1]; [1]Ctr. de GÃ©nÃ©tique Humaine, CHU de BesanÃ§on, BesanÃ§on, France, [2]Service de Neurologie PÃ©diatrique, HÃ´pital Armand Trousseau, Paris, France, [3]Dept. of Neuromuscular Disorders, UCL Inst. of Neurology, London, United Kingdom, [4]Clinical Genetics and Genomics, Royal Brompton and Harefield NHS Fndn. Trust, London, United Kingdom, [5]Service de GÃ©nÃ©tique, Hospices Civils de Lyon, Bron, France, [6]Service de GÃ©nÃ©tique Clinique, Ctr. RÃ©fÃ©rence 'DÃ©ficiences Intellectuelles de causes rares' (CRDI), Ctr. Hosp.ier Univ.ire de Rennes, Rennes, France, [7]Dept. of Genetics, PitiÃ©-SalpÃªtriÃ¨re Hosp., APHP, Sorbonne UniversitÃ©, Paris, France, [8]Service de GÃ©nÃ©tique, CHU Reims, Reims, France, [9]Necker Hosp., APHP, Reference Ctr. for Mitochondrial Diseases, Genetics Dept., Inst. Imagine, Paris, France, [10]Dept. of Neurology, Boston Children's Hosp., Boston, MA, [11]Dept. of NeuroSci.s, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Hlth., Univ. of Genoa, Genoa, Italy, [12]Pediatric Neurology and Muscular Diseases Unit, IRCCS Giannina Gaslini Inst., Genoa, Italy, [13]Dept. of Molecular Med., Univ. of Pavia, IRCCS Mondino Fndn., Pavia, Italy, [14]IWK Hlth.Ctr., Halifax, NS, Canada, [15]Dept. of Med. Sci., Univ. of Torino, Torino, Italy, [16]Clinical Genetics, CÃ´te de Nacre Univ. Hosp. Ctr., Caen, France, [17]Seoul Natl. Univ. Hospit, Seoul, Korea, Republic of, [18]CHU de Rennes Hosp SUD, Rennes CDX 2, France, [19]Med. Genetics Service and Lab. of Cytogenetics, SIC Anatomia Patologica, 'San Carlo' Hosp., Potenza, Italy, [20]Dept. of Neurology, Schneider Children's Med. Ctr. of Israel, Petah Tiqva, Israel, [21]Univ. of Melbourne, Austin Hlth.and Royal Children's Hosp., Florey Inst. and Murdoch Children's Res. Inst., Melbourne, Australia, [22]Univ California, La Jolla, CA, [23]Natl. Res. Ctr., Dokki, Cairo, Egypt., Egypt, [24]Dept. of Human NeuroSci.s, Inst. of Child and Adolescent Neuropsychiatry, Sapienza Univ. of Rome, Rome, Italy, [25]Psycho-neurological Hlth. center, Medistyle Hosp., Baku, Azerbaijan, [26]APHP-Hosp. Trousseau, 75012 PARIS, France

Disclosure Block: **C. Engel:** None.

BRAT1 bi-allelic mutations have been associated with two distinct clinical pictures: the rigidity and multifocal seizure syndrome (RMFSL) and a neurodevelopmental disorder associating cerebellar atrophy with or without seizures syndrome (NEDCAS). To our knowledge, 37 patients from 26 families have already been described with recessive variations in *BRAT1*. We report here clinical and molecular findings of 51 patients from 39 families with bi-allelic pathogenic variants in the *BRAT1* gene. Prenatal features were observed in 3 patients (3/39; 8%). Microcephaly was observed in 22 of 41 patients (54%) and was already present at birth in 6 of 45 patients (13%). Hypotonia was noted in 28 of 44 patients (64%) and hypertonia of the limbs in 23 patients (46%). Intellectual disability was noted in all patients, except one. All but this one showed developmental delay, without any psychomotor acquisition for 41% of them (21/51). Only 48% of patients (24/50) acquired walking. Ataxia was present in 41% of cases (21/51). Epilepsy was present in 51% of patients (26/51). Mean age of onset was around one year of life (358 days) ranging from in utero to 13 years old. First seizures occurred before the age of one year in 81% of cases (21/26), and in 50% of cases (13/26) in the first week of life. In the majority of cases, epilepsy was drug resistant (20/26 ; 77%). Brain MRI showed cerebral atrophy in 13 patients (13/49; 27%) and cerebellar atrophy in 34 patients (34/49; 69%). Death occurred in 18 patients (35%), before the age of 1 year for 14 of them (28%). The phenotype of our patients seems less severe than that described in the literature. This might be

explained by the difference in the representation of truncating and missense mutations. We therefore suggest that biallelic *BRAT1* variants are not associated with two distinct clinical presentations but rather with a broader phenotypic spectrum. The most severe end of this disorder, mainly seen in patients with two truncating variations, is associated with profound intellectual disability, drug-resistant epilepsy, cerebral atrophy and early death. At the other end of the spectrum, a phenotype of mild ID, cerebellar atrophy, ataxia, nystagmus and higher life expectancy is observed in patients with at least one missense variant.

PrgmNr 2590 - The pleiotropy of neurodegenerative repeat expansions in ALS

View session detail

Author Block: **J. Hengeveld**[1], A. Vajda[2], M. Heverin[2], R. L. McLaughlin[1]; [1]Smurfit Inst. of Genetics, Trinity Coll. Dublin, Dublin, Ireland, [2]Academic Unit of Neurology, Trinity BioMed. Sci. Inst., Trinity Coll. Dublin, Dublin, Ireland

Disclosure Block: **J. Hengeveld:** None.

ALS is a fatal neurodegenerative disorder which causes the death of neurons controlling voluntary muscles. ALS has no cure, and its underlying cause is mostly unknown, although a strong genetic component is known to play a role. Repeat expansions (REs) underlie more than 40 diseases, most of them affecting the nervous system. The most common neurodegenerative repeat expansions (NDREs) diseases are Huntingtonâ□□s Disease (HD), Spinocerebellar Ataxias (SCA), Frontotemporal Dementia (FTD) and Amyotrophic Lateral Sclerosis (ALS). Several REs are pleiotropic; for example, GGGCC RE in *C9orf72* is associated with FTD/ALS and CAG RE in *ATXN2* causes SCA2/ALS. Previous studies on *ATXN2* showed that harbouring intermediate-length repeat expansions are significantly associated with the risk of ALS. Therefore, pleiotropy might be common in ALS. This study aims to genotype 34 neurodegenerative genes that harbour REs, in a cohort of 1000 controls and 1000 patients from the Irish ALS bank to assess the association between expanded genotypes and ALS. The length measurement of each NDRE gene and its possible repeat expansion was done by PCR, Repeat Primed-PCR (RP-PCR), agarose gel electrophoresis and fragment length capillary electrophoresis (FLA). In an Irish population, ALS might be driven by multiple intermediate-length repeat expansion in likely 8 NDREs genes: *ATXN1*, *ATXN2*, *DIP2B*, *FRA11AC1*, *HTT*, *NUTM2B-AS1*, *PABN1* and *ZNF713*. This study will give a better understanding of ALS mechanisms. ALS is a very complex disease that might be caused by pleiotropy of multiple REs and multiple factors.

PrgmNr 2591 - Whole exome sequencing of consanguineous individuals reveals rare coding variants in known and novel genes that contribute to neurological disorders

View session detail

Author Block: **W. Koomson**[1,2], A. E. Schaffer[2], M. Gunel[1]; [1]Yale Univ., New Haven, CT, [2]Case Western Reserve Univ., Cleveland, OH
Disclosure Block: **W. Koomson:** None.
Neurologic disorders can cause structural, connective, biochemical or electrical abnormalities in the nervous system. We analyzed 1,500 patients including 70 trios and 3 quads, with neurological/neurodevelopmental disorders of varying severity including structural brain malformations from a network of 28 clinical centers in Turkey, with a high degree of consanguinity. Autosomal recessive diseases are more frequently observed in consanguineous unions, therefore homozygosity analysis allows for gene mapping of recessive mutations. Our cohort consists of 60% consanguineous families with 46% female and 54% male affected individuals. Structural brain malformations occur in 45% of patients in our cohort, often presenting with clinical phenotypes such as autism, intellectual disability, dysplasia, cerebellar malformations, epilepsy, microcephaly, lissencephaly, neurodegenerative disorders, and neurological syndromes. Whole-exome sequencing enabled us to identify known disease-causing variants in a subset of the cohort, including: C19orf12 in patients with Hallervorden-Spatz disease, CACNA1A and CACNA1B in patients with epilepsy and intellectual disability, COL18A1 in patients with Knobloch Syndrome and Frontal Pachygyria, KIAA1211L in neurodevelopmental disorders, and WDR62 in patients with microcephaly and lissencephaly. Nevertheless, the majority of disease causing variants currently remain unidentified in our cohort. In addition to our continued analysis of homozygous variants, we will also explore the role of de novo mutations, which have been shown to be increased in neurodevelopmental disorders. Finally, we will identify gene co-expression module profiles, and perform gene enrichment analysis to identify categories of genes that may have an association with disease phenotypes.

PrgmNr 2592 - Clinical implementation of RNA sequencing for Mendelian disease diagnostics

View session detail

Author Block: **V. A. Yepez**[1], M. Gusic[2], R. Kopajtich[3], J. Gagneur[4,5], H. Prokisch[6]; [1]Technical Univ. of Munich, Garching bei MÃ¼nchen, Germany, [2]Helmholtz Zentrum Munich, Munich, Germany, [3]Helmholtz Ctr. Munich, Munich, Germany, [4]Technical Univ. of Munich, Garching, Germany, [5]Inst. of Computational Biology, Munich, Germany, [6]Helmholtz Zentrum Muenchen, Neuherberg, Germany
Disclosure Block: **V.A. Yepez:** None.
Lack of functional evidence hampers variant interpretation, leaving a large proportion of cases with a suspected Mendelian disorder without genetic diagnosis after whole exome sequencing (WES). Research studies advocate to further sequence transcriptomes to directly and systematically probe gene expression defects. However, collection of additional biopsies, and establishment of lab workflows, analytical pipelines, and defined concepts in clinical interpretation of aberrant gene expression are still needed for adopting RNA-sequencing (RNA-seq) in routine diagnostics. We implemented an automated RNA-seq protocol and developed a computational workflow, DROP, with which we analyzed skin fibroblasts of more than 300 individuals with a suspected mitochondrial disease which previously underwent WES. We detected on average 12,500 genes per sample including around 60% disease genes - a coverage substantially higher than with whole blood, supporting the use of skin biopsies. We prioritized genes demonstrating aberrant expression using OUTRIDER, aberrant splicing using FRASER, or mono-allelic expression. The pipeline required less than one week from sample preparation to result reporting and provided a median of eight disease-associated genes per patient for inspection. A genetic diagnosis was established for 16% of the 205 WES-inconclusive cases. Detection of aberrant expression was a major contributor to diagnosis including instances of 50% reduction, which, together with mono-allelic expression, allowed for the diagnosis of dominant disorders caused by haploinsufficiency. Moreover, calling aberrant splicing and variants from RNA-seq data enabled detecting and validating splice-disrupting variants, of which the majority fell outside WES-covered regions. We share DROP, which besides integrating all analysis steps, it also includes quality control, can leverage parallel computing infrastructures and generates browsable web page reports. Together, these results show that streamlined experimental and computational processes can accelerate the implementation of RNA-seq in routine diagnostics.
Preprint at https://doi.org/10.1101/2021.04.01.21254633

PrgmNr 2594 - Dietary nutritional recommendations in patients with phenylketonuria in Ecuador

View session detail

Author Block: **A. Campodonico**[1], E. Haro[1], B. Bahamonde[1], J. Pozo[1], V. I. Romero[2]; [1]USFQ, Quito, Ecuador, [2]Univ. San Francisco de Quito, Quito, Ecuador
Disclosure Block: **A. Campodonico:** None.
Phenylketonuria (PKU) is an autosomal recessive disease due to an enzymatic deficit to convert the phenylalanine into tyrosine, a crucial component in neurodevelopment from the fetal stage. Without treatment it results in irreversible intellectual disability. The Ecuadorian neonatal screening tests for PKU, however it does not provide nutritional recommendations or meal guides including available products. Our study aims to assess the PKU individual nutritional status, develop nutritional guidelines with Ecuadorian products and educate the families with them.
First, we performed a 24-hour diet reminder in patients diagnosed with PKU. The test recorded a day of complete feeding, allowing a real starting point in the amount of phenylalanine consumed daily. The results indicated higher levels of phenylalanine than recommended, since they do not have a personalized nutrition guide.
Second, we performed a systematic literary review in English and Spanish of the nutritional databases available in Latin America, however we didnÂ´t find any regional guideline. Therefore, we used nutritional guidelines from North America and Europe to develop a Latin American nutritional recommendation guide with products from our region.
In conclusion, we identified that Ecuadorian patients with PKU do not have a nutritional counseling for their condition and are consuming higher levels of phenylalanine than recommended. We did not find Latin American nutritional guidelines, so we are developing one, with products from our regions.

PrgmNr 2595 - Dysosteosclerosis in an individual harboring a homozygous nonsense variant in *TNFRSF11A*: further considerations on genotype-phenotype correlation

View session detail

Author Block: **M. Castro**[1], B. D. W. Linnenkamp[1], C. R. Gama[2], T. M. Ali[3], L. A. N. Oliveira[1], R. S. Honjo[1], G. L. Yamamoto[4], C. A. Kim[1], D. R. Bertola[1]; [1]Inst. da Crianca HCFMUSP, Sao Paulo, Brazil, [2]Inst. Central HCFMUSP, Sao Paulo, Brazil, [3]Inst. de Biociencias USP, Sao Paulo, Brazil, [4]Inst. da Crianca HCFMUSP, SÃ£o Paulo, Brazil
Disclosure Block: **M. Castro:** None.
Dysosteosclerosis and osteopetrosis are both sclerosing skeletal dysplasias with overlaping radiographic signs and classified in the group 23 of the nosology of genetic skeletal disorders. They are classically differentiated by the findings of platyspondyly and increased submetaphyseal radiolucency, associated only to dysosteosclerosis, and by the presence of bone marrow failure, a hallmark complication of osteopetrosis. Although they present genetic heterogeneity with variants in distinctive genes, recently they converged, as both could be caused by biallelic variants in *TNFRSF11A*. A previously proposed hypothesis for genotype-phenotype correlation states that variants leading to synthesis of truncated or elongated RANK proteins lead to the milder dysosteosclerosis phenotype, while total loss of expression of all isoforms could be associated to osteopetrosis. We report on an adult individual with clinical and radiological findings compatible with dysosteosclerosis and absence of hematological abnormalities, harboring the nonsense variant NM_003839.4(*TNFRSF11A*):c.1325G>A:p.(Trp442*) in a homozygous state. This variant creates a stop codon only in isoforms 1 and 2, predicted to be degraded by nonsense-mediated mRNA decay, and does not affect splice site predictions. Thus, isoforms 3, 4 and 5 are probably not affected. We have noticed that in all patients with dysosteosclerosis reported so far, there is total or partial preservation of isoforms 4 and 5. Therefore, our case illustrates that only the maintenance of the expression of these two isoforms (not necessarily the presence of truncated or elongated proteins) could explain the lack of hematologic abnormalities, or at most, cause a mild compromise.

PrgmNr 2596 - Expanding the clinical spectrum of *NDUFAF3*-related disorders: Prolonged survival and unusual course in a patient with compound heterozygous *NDUFAF3* variants

View session detail

Author Block: **A. van der Ven**[1], I. Wente[2], R. Feichtinger[3], D. Weiss[4], T. Bierhals[1], K. Tsiakas[4], H. Prokisch[5,6], R. Santer[4], R. Kopajtich[5], J. A. Mayr[3], I. Wittig[2], M. Hempel[1]; [1]Dept. of Human Genetics, Univ. Med. Ctr. Hamburg-Eppendorf, Hamburg, Germany, [2]Inst. of Cardiovascular Physiology, Goethe Univ., Frankfurt a.M., Germany, [3]Univ. Children's Hosp., Salzburger Landeskliniken (SALK) and Paracelsus Med. Univ. (PMU), Salzburg, Austria, [4]Dept. of Pediatrics, Univ. Med. Ctr. Hamburg-Eppendorf, Hamburg, Germany, [5]Helmholtz Zentrum Muenchen, Neuherberg, Germany, [6]Inst. of Human Genetics, Klinikum Rechts der Isar, TUM Munich, Munich, Germany
Disclosure Block: **A. van der Ven:** None.
Mitochondrial disorders most frequently originate from impaired oxidative phosphorylation (OXPHOS) activity. Mitochondrial complex I (CI) is the largest OXPHOS component. It is composed of three functional modules: the Q-, the P- and the N-module. NDUFAF3 (NADH: Ubiquinone Oxidoreductase Complex Assembly Factor-3) is one of 13 proteins known to aid the correct assembly of this large multienzyme complex. Recent investigations have advanced the understanding of NDUFAF3 for CI biogenesis, pinpointing it to the step were the Q- and the P-module are joined. Recessive disease-causing variants in *NDUFAF3* have previously been described as the cause of a distinct mitochondrial disease condition (OMIM # 618240). The seven patients reported to date exhibited severe neurologic symptoms and lactic acidosis, followed by death within the first two years of age in six cases. We present a 10-year-old patient with developmental delay, exercise intolerance, dystonia, basal ganglia abnormalities, and elevated lactate concentration in blood. Trio-exome sequencing revealed a paternally inherited splice site variant and a maternally inherited missense variant in *NDUFAF3*, classified as pathogenic and likely pathogenic, respectively, according to ACMG criteria. Spectrophotometric analysis of fibroblast-derived mitochondria demonstrated isolated and relatively mild reduction of CI in our patient. Complexome analysis revealed undetectable levels of NDUFAF3 and a severely decreased amount of fully assembled CI. Furthermore, an unusual occurrence of the assembly factors ACAD9 and ECSIT was noted on fully assembled complex I. With this untypical presentation of a patient with onset of symptoms during early childhood and comparatively attenuated course, we provide further insight into the phenotypic spectrum of NDUFAF3-related mitochondrial disease. The relatively mild phenotype of our patient is likely a result of the paternally inherited splice-site variant. This variant is relevant for only one of the four coding *NDUFAF3* transcript-isoforms, accounting for approximately 25% of NDUFAF3 expression across human tissues. The data obtained via complexome analysis are in line with the previously defined role of NDUFAF3 for CI biogenesis. Beyond that, the abnormal occurrence of the assembly factors ACAD9 and ECSIT additionally indicates an incomplete maturation of CI in our patient.

PrgmNr 2597 - Gut microbiome affects human bone mass, microarchitecture, and strength: The Framingham Study

View session detail

Author Block: **P. Okoro**[1], E. Orwoll[2], M. Bouxsein[3], S. Sahni[1], D. P. Kiel[1]; [1]Hebrew SeniorLife, Boston, MA, [2]Oregon Hlth.& Sci. Univ, Portland, OR, [3]Beth Israel Deaconess Med. Ctr., Boston, MA
Disclosure Block: **P. Okoro:** None.
The gut microbiome has been shown to affect the inflammatory environment through effects on the T-cell landscape, which influence the production of soluble immune mediators and inflammatory cytokines that stimulate osteoclastogenesis and bone loss in mice. However, there is a lack of large human studies of the gut microbiome and skeletal health. Thus, we explored the relationship between the gut microbiome and measures of volumetric bone mineral density (vBMD), microarchitecture, and strength in humans.
We used 16S rRNA amplicon sequencing and high resolution peripheral quantitative computed tomography (HR-pQCT) phenotypes from men and women in the Framingham Study (n=1310). The 16S data were processed into operational taxonomic units (OTUs) and matched to the Greengenes 16S database. After QC and sequencing batch correction, the resulting OTU counts were converted into relative abundances. HR-pQCT scans of the distal radius and tibia generated volumetric density and architecture of cortical (porosity & thickness) and trabecular bone (thickness & number), and biomechanical estimates of failure load.
To probe for the association between microbial genera and bone measures using generalized linear model, we normalized and transformed the OTU abundances, and adjusted the bone measures for age, sex, BMI, smoking and diabetes status. At FDR â¤ 0.05, 10 unique genera were associated with one or more bone measures. We found interesting patterns of associations across the skeletal phenotypes. First, 75% of the associations were at the tibia; in the majority of these, greater abundance was associated with worse bone measures. The remainder (25%) of the associations at the radius were equally bad for radial skeletal site measures. Additionally, all bone-taxa associations at the radius were observed only with trabecular measures while those at the tibia were observed mostly with cortical measures. We did observe two microbes, *Eubacterium dolichum* and *Coprobacillus cateniformis*, for which greater abundances were associated with lower measures of bone density, worse micro-architecture (lower cortical thickness) and bone strength (reduced failure load).
In summary, these early results from a moderately large cohort study show evidence of significant relationships between gut microbiota and bone measures that differ between cortical and trabecular bone and in weight bearing tibial and non-weight bearing radial sites. Several microbes demonstrated deleterious associations across measures of density, microarchitecture and strength, thus supporting the importance of the gut microbiota in skeletal homeostasis and health.

PrgmNr 2598 - Novel *RUNX2* gene variation associated with cleidocranial dysplasia-like phenotype

View session detail

Author Block: **J. Aponte**[1], J. Pascual[2], Y. Mojica[3], C. I. Hernandez[4], S. Carlo[5], A. S. Cornier[6]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Ponce Hlth.Sci. Univ. Sch. of Med., Ponce, Puerto Rico, [3]Univ. of Puerto Rico Rio Piedras Campus, San Juan, Puerto Rico, [4]Natl. Children's Hosp., Pediatrics Dept., Washington D.C., DC, [5]Ponce Hlth.Sci. Univ., Cabo Rojo, PR, [6]San Jorge Children's Hosp. / Ponce Hlth.Sci. Univ., San Juan, PR

Disclosure Block: **J. Aponte:** None.

The *RUNX2* gene has been recognized to play an important role in the formation of bone, cartilage, and teeth, while several variations have been associated with the development of cleidocranial dysplasia (CCD). CCD affects the development of the bones, skull, and teeth. Signs and symptoms include underdeveloped or absent clavicles, dental abnormalities, and delayed closing of the fontanelles. Other symptoms may include osteopenia, hearing loss, bone abnormalities of the hands, and recurrent sinus and ear infections. People with CCD may develop curvature of the spine (scoliosis), osteoporosis, and may be shorter than average in stature. *RUNX2* is inherited in an autosomal dominant pattern. DNA sequence analysis of an 18-year-old patient with severe progressive scoliosis, webbed neck, clavicle dysplasia, fragile teeth and brachydactyly revealed a novel variation within one copy of the *RUNX2* gene (c.228_260dup) which results in the insertion of 11 amino acids to the RUNX2 protein (p.Ala79_Ala89dup). The presence of this genetic variation, along with the patient's clinical presentation, suggests a novel phenotype-genotype association, characterized by the appearance of both typical and unusual signs and symptoms of cleidocranial dysplasia in this individual.

PrgmNr 2599 - Novel variants of PAH gene in Ecuadorian phenylketonuria patients: Do we need new molecular panels for metabolic disorders for Latin American populations?

View session detail

Author Block: **E. Haro**[1], A. CampodÃ³nico[1], B. Bahamonde[1], J. Pozo[1], V. I. Romero[2]; [1]USFQ, Quito, Ecuador, [2]Univ. San Francisco de Quito, Quito, Ecuador
Disclosure Block: **E. Haro:** None.
The heterogeneity of the molecular variants of the *phenylalanine hydroxylase (PHE)* gene results in a spectrum of enzyme deficiency and mild to severe phenotypes. PKU is part of the nationwide neonatal screening, however molecular analysis is not covered or studied.
We identified the molecular variant on previously diagnosed individuals with PKU and report novel variants. Ecuador is a country with an admixed population between African, European, and Native American groups. Therefore, we found both common and novel molecular variants in our patients linked to these populations in literature and databases. Most of genetic databases have limited participants from minorities, we concluded that the novel variants arise from Indigenous populations of the region
Our findings highlight the importance of developing population-related molecular panels in Latin American. We correlate the molecular variants and the complexity of nutritional treatment, demonstrating the importance to offer molecular analysis to patients with PKU to optimize their management.

PrgmNr 2600 - Sensorineural hearing loss is an early indicator of mitochondrial disease

View session detail

Author Block: **A. Koleilat**, G. L. Poling, S. C. Ekker, V. S. Pureza, L. A. Schimmenti, L. Hasadsri; Mayo Clinic, Rochester, MN
Disclosure Block: **A. Koleilat:** None.
Disorders of energy metabolism are multisystem conditions characterized by dysfunctional mitochondria caused by genetic mutations in either mitochondrial DNA (mtDNA) or nuclear DNA (nDNA) and are collectively known as Mitochondrial Disease (MD). Tissue distribution (replicative segregation) and heteroplasmy contribute to the multiple system nature of MD and impact age of clinical presentation. Sensorineural hearing loss (SNHL) occurs in MD and can range in severity and age of onset. It is estimated that over half of patients with MD will develop some hearing impairment and of those individuals, more than 50% experienced hearing loss prior to age 40. In this retrospective study, we evaluated patients referred to a large reference laboratory who were identified to have diagnostic testing supportive of MD and reviewed medical records for the presence, type, and age of onset of SNHL. The goal of this study was to determine if a genetic investigation occurred at the onset of hearing loss would an earlier diagnosis of MD occurred, and furthermore, if hearing loss can be used as an early indicator for MD. Of the 50 total patients identified with a confirmed molecular MD diagnosis, 42% were diagnosed between the ages of 21- 40. Within this cohort 72% had a missense variant in mtDNA, 16% had a deletion in mtDNA and12% had a missense pathogenic variant in nDNA. Medical records were searched for evidence of a SNHL diagnosis prior to a mitochondrial disease diagnosis. Thirty-six percent of this cohort presented with SNHL prior to the diagnosis of MD. Of the 36, audiometric data (1-8 kHz) was available for 18 individuals. In all 18, mild to moderate hearing loss at lower frequencies and moderate hearing loss at 2 kHz and higher was identified. In individuals where a specific age was recorded for the onset of hearing loss, there was an average difference of 15 years between discovery of hearing loss and the diagnosis of mitochondrial disease. In conclusion, we discovered that over one-third of individuals with a molecular MD diagnosis presented with SNHL, on average, over a decade prior to being tested for and diagnosed with MD. This observation supports the need for a clinical genetics evaluation that includes consideration of MD in the work up of individuals who present with early adult onset of SNHL This suggests that SNHL could be an early indicator of MD in a subset of patients.

PrgmNr 2601 - The role of seryl-tRNA synthetase in mitochondrial biology and disease

View session detail

Author Block: **C. Del Greco**[1], J. Kitzman[2], A. Antonellis[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ MICHIGAN, Ann Arbor, MI
Disclosure Block: **C. Del Greco:** None.
Aminoacyl-tRNA synthetases (ARSs) are essential, ubiquitously expressed enzymes that charge tRNA molecules with amino acids in the cytoplasm and mitochondria. All 37 human ARSs are nuclear-encoded: 17 function in the cytoplasm, 17 function in the mitochondria, and three function in both cell compartments. Importantly, all 17 mitochondrial ARSs have been implicated in human recessive disease with a broad range of clinical phenotypes that often affect tissues with high energy demands. Mitochondrial seryl-tRNA synthetase (*SARS2*) has been implicated in disease phenotypes ranging from progressive spastic paresis to HUPRA (hyperuricemia, pulmonary hypertension, renal failure in infancy, and alkalosis) syndrome, but there is no explanation for this variability. Additionally, the number of patient variants, all of which map to the catalytic domain of the protein, is limited, and the full allelic spectrum of potentially pathogenic *SARS2* mutations remains unknown. To expand the allelic spectrum of *SARS2*-related disease, I will perform a massively parallel assessment of *SARS2* exons encoding key functional domains by performing saturation genome editing on haploid (Hap1) human cells. First, I will develop a Hap1 cell line that harbors an integrated, doxycycline-inducible copy of *SARS2*; the *SARS2* transgene will be engineered to be resistant to subsequently used gRNAs. Second, I will engineer all possible *SARS2* missense and nonsense mutations in 8 exons of interest; this will result in a total of 2934 *SARS2* variants studied. Third, I will use CRISPR/Cas9 genome editing to introduce the mutation library into a population of Hap1 cells and will determine mutation integration frequencies in the presence of doxycycline (*i.e.*, during expression of the wild-type *SARS2* transgene). Finally, I will remove doxycycline treatment, allow the cells to grow for 4 days, and sequence DNA from the resulting population. Variants that are reduced in frequency after doxycycline removal will represent loss-of-function variants that impair cell viability. In parallel, I will model known patient variants in Hap1 cells and investigate their effects on mitochondrial protein translation and oxygen consumption to correlate mitochondrial function with clinical severity. Here, I will present my plans to study *SARS2* variants and our unpublished, preliminary data. This study will inform our understanding of essential regions of SARS2, broaden our understanding of the allelic heterogeneity of pathogenic *SARS2* alleles to assist in assessing newly identified patient mutations, and deepen our understanding of how mitochondrial ARSs lead to diverse disease phenotypes.

PrgmNr 2602 - A digital health approach to rare disease diagnosis; identifying patients with feature combinations warranting further evaluation for a possible diagnosis of 22q11.2 deletion syndrome from UK primary care electronic health records

View session detail

Author Block: **L. Menzies**, O. Buendia, W. Evans; Mendelian, London, United Kingdom
Disclosure Block: **L. Menzies:** Consultant/Consulting Fees/Other Remuneration; Mendelian Ltd.
Introduction: 22q11.2 deletion syndrome is a rare disease with multi-systemic features including congenital cardiac malformations, velopharyngeal insufficiency, parathyroid hypoplasia and intellectual disability (prevalence of ~25/100,000). Rare disease diagnosis is challenging due to disease complexity and low physician awareness. The 2021 UK Rare Diseases Framework highlights a global need for faster diagnosis, to improve clinical outcomes.
Methods: We developed a digital health tool to facilitate earlier diagnosis of rare diseases. The tool scans pseudo-anonymised Electronic Health Records (EHR) to highlight patients for further investigation for a given syndrome. Peer-reviewed disease criteria are mapped to the appropriate SNOMED CT codes to create a digital criteria algorithm. An algorithm for 22q11.2 deletion syndrome was derived from a published diagnostic criterion. The tool identifies patients who breach a defined threshold depending on the diagnostic significance of SNOMED CT codes. This digital algorithm was applied to 501,188 primary care EHR. Matched cases were subject to internal review to consider appropriateness for further evaluation. EHR of patients that matched the 22q11.2 algorithm and those already diagnosed with 22q11.2 deletion syndrome were reviewed.
Results: 25 patientsâ EHR had a diagnosis of 22q11.2 deletion syndrome. The digitised criteria algorithm identified 5 of these patients; 3 of whom matched the digitised criteria in advance of the diagnostic code being assigned. For example, 1 patient met the digitised criteria threshold for suspected diagnosis 11 years prior to recorded diagnosis. 112 undiagnosed patientsâ EHR also matched the digitised criteria; 102 of these were deemed appropriate for further evaluation after an internal manual review.
Conclusions: This digital approach has the potential to identify patients with a possible diagnosis of 22q11.2 deletion. Improved (earlier) diagnosis has important implications for clinical care, for example enabling surveillance for disease complications and genetic counselling. Further prospective studies are planned to evaluate this digital approach and its implementation as an adjunctive tool in primary care.

PrgmNr 2603 - Association of missense variants in the transthyretin *TTR* gene with hereditary transthyretin-mediated hATTR amyloidosis in 454,782 UK Biobank whole exome sequences

View session detail

Author Block: **M. Plekan**[1], R. Hoffing[1], L. D. Ward[2], A. Deaton[1], A. Flynn-Carroll[1], C. Willis[1], G. Hinkle[1], P. Nioi[2], M. M. Parker[2]; [1]Alnylam Pharmaceuticals, Cambridge, MA, [2]Alnylam, Cambridge, MA
Disclosure Block: **M. Plekan:** Salary/Employment; Alnylam Pharmaceuticals.
ObjectiveOver 100 distinct missense variants in the TTR gene are known to cause hereditary transthyretin-mediated (hATTR) amyloidosis, a progressively debilitating life-threatening disease. We aimed to characterize missense TTR variants in a cohort of 454,782 exomes and test their association with hATTR amyloidosis. MethodsMissense TTR variants were identified in 454,782 exomes from the UK Biobank, a population-based prospective study where patients are followed to age 68 on average. Variants were classified as pathogenic, likely pathogenic, benign, conflicting interpretation, uncertain significance or novel based on CLINVAR. â□□Novelâ□□ indicates the variant had not been previously assessed for pathogenicity. In carriers of known pathogenic variants, the prevalence of an hATTR amyloidosis diagnosis and common hATTR symptoms (polyneuropathy, carpal tunnel syndrome, heart failure and cardiomyopathy) were characterized. In carriers of variants with uncharacterized/uncertain pathogenicity, we assessed diagnoses for evidence of hATTR amyloidosis.Results We identified 84 missense TTR variants including 10 pathogenic variants (N=423), 20 variants of uncertain or conflicting interpretation (N=740), and 51 novel variants (N=125). The most prevalent pathogenic variants were V122I (N=358), V30M (N=25), T60A (N=19) and I68L (N=12). Ten of 423 pathogenic variant carriers (2.4%) had an amyloidosis diagnosis (ICD10 code â□□E85â□□). Of pathogenic variants with >1 carrier, the mutation with the highest frequency of amyloidosis diagnosis was T60A (15%) followed by V122I (2%). By age 75, 12.1% (95% CI = 6.7-17.2%) of carriers had at least one common symptom of hATTR amyloidosis. Phenome-wide association analysis of the common missense variant Gly6Ser (rs1800458, MAF=8%, N=64,353) in the white British population (N=363,973) revealed no significant associations supporting its benign CLINVAR designation. Analysis of carriers of missense TTR variants with uncharacterized pathogenicity revealed no evidence of novel pathogenic hATTR amyloidosis variants. Lastly we identified 15 heterozygous carriers of high confidence predicted loss of function mutations, none of whom had an amyloidosis diagnosis.ConclusionsWe identified 84 missense TTR variants in 454,782 UKBB exomes, including 10 known pathogenic variants and 71 variants of unknown pathogenicity. Despite a low frequency of amyloidosis diagnosis among pathogenic variant carriers, many have diagnoses of common hATTR symptoms, suggesting an underdiagnosis of disease. No evidence of additional hATTR amyloidosis pathogenic variants among missense variants of unknown pathogenicity was found.

PrgmNr 2605 - Increasing the diagnostic yield through post-report reanalysis of exome data in subjects with immune disorders

View session detail

Author Block: **M. A. Walkiewicz**[1], M. Similuk[2], B. Seifert[3], J. Yan[4], A. Borges[1], M. Setzer[5], C. Jodarski[6], M. Kamen[4], D. Veltri[1], S. Holland[1], R. Ghosh[2]; [1]NIAID/NIH, Bethesda, MD, [2]NIH, Bethesda, MD, [3]Natl. Inst. of Allergy and Infectious Diseases (NIAID), Bethesda, MD, [4]Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, [5]Med. Sci. and Computing/NIAID, Bethesda, MD, [6]
Disclosure Block: **M.A. Walkiewicz:** None.
Background & Rationale:The rate of molecular diagnosis within immunology by exome sequencing ranges between 15%-79% depending on the population and technical approach. Several studies have shown that periodic reanalysis yields additional diagnoses. Initial analysis of the first 1000 probands of our cohort of immunology patients referred to the National Institute of Allergy and Infectious Diseases (NIAID) Centralized Sequencing Program (CSP) yielded 341 molecular diagnoses among 314 probands (31.4%). Here, we present our reanalysis of the first 1000 probands. Methods:Research-based exome sequencing was performed on 1000 study probands. In addition, chromosomal microarray was performed in 374 of these probands. The data was then analyzed, and relevant findings were reported to the subjects. Semi-automated reanalysis of the data was performed between 1-3 years after the initial review and focused on new gene-disease associations, new HGMD-cataloged variants present in our database, as well as updates in clinical presentation or inclusion of additional relatives in the analysis. Results:Reanalysis yielded 21 new molecular diagnoses in 20 probands (2.0%), six of whom already had at least one molecular diagnosis. In this group, there were 23 variants which now constituted or contributed to a molecular diagnosis. Nine diagnoses were made from newly described gene-disease associations. Among these, 3 patients with variants in *SASH3* were initially identified through the CSP and published in collaboration with NIAID investigators, one patient with a truncating variant in *IL6ST* contributed to the original research report, and one diagnosis came from a new CNV-disease association. In addition, 11 molecular diagnoses were made from re-assessment of the variant due to new publications, data on relatives or updated clinical history. Lastly, two molecular diagnoses were made from the updated ACMG incidental findings gene list. Conclusion:In summary, our reanalysis of 1000 probands yielded an additional 21 diagnoses, increasing our diagnostic rate to 32.8% (328 probands with 362 molecular diagnoses). With the fast pace of new gene-disease discoveries, incorporating newly reported genes into reanalysis is important in maintaining the most up to date analysis pipeline. Periodic review of new literature of variants in disease genes and inclusion of family members for segregation studies proved to be important factors contributing to new diagnoses. Additionally, reanalysis of CNV data should be considered as the contribution of CNVs to immune disorders is not fully understood. Lastly, scalability should remain a key consideration for all re-analysis methods.

PrgmNr 2606 - Pediatric onset cardiomyopathy secondary to *TNNC1* heterozygous and homozygous mutations: a French cohort study

View session detail

Author Block: **M. WÃ©ber**[1], S. Falcon-Eicher[2], G. Millat[3,4], D. Khraiche[5], Y. Alembik[6], S. El Chehadeh[6], G. Delavilleon[7,8], A-C. Brehin[9], J. Albuisson[10,11], J-C. Eicher[12], C. Thauvin-Robinet[13,14], P. Charron[15,16], P. Richard[16,17], L. Faivre[1,14]; [1]Ctr. de RÃ©fÃ©rence Anomalies du DÃ©veloppement et Syndromes Malformatifs, FHU TRANSLAD, HÃ´pital d'Enfants, CHU Dijon, Dijon, France, [2]UnitÃ© de Cardiologie PÃ©diatrique et CongÃ©nitale Adulte, HÃ´pital d'Enfants, CHU Dijon, Dijon, France, [3]Laboratoire de CardiogÃ©nÃ©tique MolÃ©culaire, Ctr. de Biologie et Pathologie Est, Hospices Civils de Lyon, Bron, France, [4]INMG - Inst. NeuroMyoGÃ¨ne, INSERM U1217, CNRS UMR 5310, UniversitÃ© Claude Bernard Lyon 1, Lyon, France, [5]Ctr. de RÃ©fÃ©rence Malformations Cardiaques CongÃ©nitales Complexes â M3C, Cardiologie CongÃ©nitale et PÃ©diatrique, HÃ´pital Univ.ire Necker-Enfants Malades, AP-HP, Paris, France, [6]Service de GÃ©nÃ©tique MÃ©dicale, Inst. de GÃ©nÃ©tique MÃ©dicale dâAlsace, CHU Hautepierre, Strasbourg, France, [7]PÃ©diatrie SpÃ©cialisÃ©e - Malformations Cardiaques, DÃ©partement de PÃ©diatrie, HÃ´pital Arnaud de Villeneuve, CHU de Montpellier, Montpellier, France, [8]UnitÃ© dâEvaluation, de RÃ©adaptation et de RÃ©habilitation en CardiopÃ©diatrie, Inst. Saint-Pierre, Palavas-Les-Flots, France, [9]Ctr. de RÃ©fÃ©rence Anomalies du DÃ©veloppement et Syndromes Malformatifs, FHU G4 GÃ©nomique, INSERM UMR U1245, CHU de Rouen, Rouen, France, [10]Plateforme de Transfert en Biologie CancÃ©rologique, GIMI - Genetic and Immunology Med. Inst., Ctr. Georges-FranÃ§ois Leclerc, Dijon, France, [11]INSERM UMR1231, UniversitÃ© de Bourgogne, Dijon, France, [12]Service de Cardiologie 2: Rythmologie et Insuffisance Cardiaque, HÃ´pital FranÃ§ois Mitterrand, Dijon, France, [13]Ctr. de RÃ©fÃ©rence DÃ©ficiences Intellectuelles de Causes Rares, FHU TRANSLAD, HÃ´pital d'Enfants, CHU Dijon, Dijon, France, [14]GAD - GÃ©nÃ©tique des Anomalies du DÃ©veloppement, INSERM UMR1231, UniversitÃ© de Bourgogne, Dijon, France, [15]Ctr. de RÃ©fÃ©rence des Maladies Cardiaques HÃ©rÃ©ditaires, DÃ©partements de GÃ©nÃ©tique et de cardiologie, Groupe Hosp.ier PitiÃ© SalpÃªtriÃ¨re, AP-HP, Paris, France, [16]ICAN - Inst. for Cardiometabolism and Nutrition, UnitÃ© de Recherche sur les Maladies Cardiovasculaires et MÃ©taboliques, INSERM UMRS 1166, Sorbonne UniversitÃ©, Paris, France, [17]Ctr. de GÃ©nÃ©tique MolÃ©culaire et Chromosomique, UF de CardiogÃ©nÃ©tique et MyogÃ©nÃ©tique MolÃ©culaire et Cellulaire, Groupe Hosp.ier PitiÃ© SalpÃªtriÃ¨re, AP-HP, Paris, France

Disclosure Block: **M. WÃ©ber:** None.

Background: *TNNC1* encodes for cardiac troponin C, a subunit of the troponin complex which constitutes thin filaments and regulates muscle contraction. It is known to be associated with a rare form of autosomal dominant cardiomyopathy. Nevertheless, after reviewing the published case reports, about one third of patients had pediatric onset cardiomyopathies.

Material and methods: We diagnosed a *TNNC1* pathogenic variant in a patient with dilated cardiomyopathy detected at one-month-old. Therefore, we set up a national collaboration call with the five French referent molecular laboratories for hereditary cardiomyopathies. We requested the number of patients diagnosed with pediatric onset cardiomyopathy out of their cohort of *TNNC1* mutated patients. Genetic analysis was performed by targeted Next-Generation Sequencing and/or Sanger sequencing.

Results: A cohort of ten children out of thirty-seven patients with *TCCN1* variants (27%) was isolated in France. Genetic investigations identified seven patients with a heterozygous variant in *TNNC1* and three patients with homozygous variants, including one brother and one sister. All were missense mutations, two of which were newly described: p.(Phe27Val) and p.(Pro54Arg) and four previously published: p.(Ala8Val), p.(Asp149Asn), p.(Asp88Asn) and p.(Asp25Asn). The three homozygous probands had the same p.(Ala8Val) mutation. Clinical assessment was only available for eight

patients. It showed that first signs mostly began before the age of one year, ranging from antenatal onset to 8-year-old. Predominant symptoms consisted of dilated (6/8) or hypertrophic cardiomyopathy (3/8). Additional conditions comprised restrictive cardiomyopathy (2/8), ventricular (2/8) or atrial septal defect (1/8) and myocardial bridge (1/8). In the majority of cases, clinical presentation was severe. It rapidly led to transplantation in one homozygous proband at 2 years old and in one heterozygous proband at 4 years old. Death occurred in three children at ages 1, 6 and 12 months, including the other two homozygous patients.
Conclusion: Neonatal or early pediatric onset cardiac phenotypes seem to be found in one third of patients with *TNNC1* mutations. Homozygous or compound heterozygous variants appeared to be responsible, in absence of transplantation, for a lethal outcome. Further reports and delineation of the genotype-phenotype correlations of this gene should improve disease management and genetic counselling.

PrgmNr 2607 - Rare variant burden influences the rate of disease progression in Polycystic Kidney Disease

View session detail

Author Block: **K. Benson**[1], E. Elhassan[2], S. F. Madden[1], S. L. Murray[3], O. Sadeghi-Alavijeh[4], J. Carmichael[4], D. Gale[4,5], C. Kennedy[6], M. Griffin[7,8], L. Casserly[9], B. Moloney[9], P. Oâ□□Hara[9], J. Abboud[10], K. Cronin[10], D. Oâ□□Sullivan[10], A. Mallawaarachchi[11], I. Capelli[12], G. LaManna[12], C. Graziano[13], R. SchÃ¶nauer[14], J. Halbritter[14], G. L. Cavalleri[15], P. Conlon[1,3]; [1]Royal Coll. of Surgeons in Ireland, Dublin, Ireland, [2]Dept. of Nephrology, Beaumont Hosp., Dublin, Ireland, [3]Beaumont Hosp., Dublin, Ireland, [4]Dept. of Renal Med., Univ. Coll. London, London, United Kingdom, [5]Genomics England, Queen Mary Univ. of London, London, United Kingdom, [6]Kingston Gen. Hosp. & Queenâ□□s Univ., Kingston, ON, Canada, [7]Regenerative Med. Inst. (REMEDI) at CÃ□RAM Ctr. for Res. in Med. Devices, Sch. of Med., Natl. Univ. of Ireland Galway, Galway, Ireland, [8]Nephrology Dept., Galway Univ. Hosp., Saolta Univ. Hlth.care Group, Galway, Ireland, [9]Dept. of Renal Med., Univ. Hosp. Limerick, Limerick, Ireland, [10]Dept. of Med., Beaumont Hosp., Dublin, Ireland, [11]Div. of Genomics and Epigenetics, Garvan Inst. of Med. Res., Sydney, Australia, [12]Dept. of Experimental Diagnostic and Specialty Med. (DIMES), Nephrology, Dialysis and Renal Transplant Unit, S. Orsola Hosp., Univ. of Bologna, Bologna, Italy, [13]Unit of Med. Genetics, S. Orsola-Malpighi Hosp., Bologna, Italy, [14]Div. of Nephrology, Dept. of Endocrinology, Nephrology, and Rheumatology, Univ. of Leipzig Med. Ctr., Leipzig, Germany, [15]RCSI, Dublin, Ireland

Disclosure Block: **K. Benson:** None.

Introduction: Autosomal dominant polycystic kidney disease (ADPKD) is caused primarily by variants in *PKD1*, and *PKD2.* Disease severity ranges from *in-utero* onset to preserved kidney function into old age. It is well established that the type of diagnostic ADPKD variant can influence disease severity and variants have been identified which modify disease severity or can cause autosomal recessive disease. Here we demonstrate, using robust statistical analyses, that rare variant burden in the gene *PKD1* contributes to ADPKD phenotypic variability.

Methods: Patients (n=449) with an established genetic diagnosis of ADPKD due to a pathogenic variant in *PKD1* or *PKD2* were recruited from five international centres (Dublin Ireland, Genomics England/ Royal Free Hospital London UK, Leipzig Germany, Bologna Spain, Sydney Australia). The association between the presence of rare, additional *PKD1* variants and age at end-stage renal disease (ESRD) was tested using a Cox mixed-effect regression model.

Results: The presence of non-diagnostic, rare, predicted damaging *PKD1* variants was associated with a lower age at ESRD, in patients with an established genetic diagnosis (hazard ratio: 1.62 (1.15-2.29), p=5.2x10^{-3}).

Conclusions: Rare, additional predicted damaging *PKD1* variants impact disease severity in ADPKD. Patients with an additional, rare, predicted damaging *PKD1* variant reached ESRD a median of 3-7 years earlier than those lacking such qualifying variants. These findings have important implications for patient counselling and the assessment of variant pathogenicity in clinical genetics services.

Acknowledgement: This research was made possible through access to the data and findings generated by the 100,000 Genomes Project; http://www.genomicsengland.co.uk.

PrgmNr 2608 - 13q Deletion Syndrome involving RB1: characterization of a new minimal critical region for psychomotor delay

View session detail

Author Block: **F. Privitera**[1], A. Calonaci[2], G. Doddato[1], F. T. Papa[1], M. Baldassarri[1], A. Pinto[3], F. Mari[4], I. Longo[3], M. Caini[2], D. Galimberti[2], T. Hadjistilianou[5], S. De Francesco[6], A. Renieri[4], F. Ariani[4]; [1]Med. Genetics Univ. of Siena Italy; Med Biotech Hub and Competence Ctr. Dept. of Med. Biotechnologies Univ. of Siena Italy, Siena, Italy, [2]Unit of Pediatrics Dept. of Maternal Newborn and Child Hlth.Azienda Ospedaliera Univ.ria Senese Policlinico 'Santa Maria alle Scotte' Siena Italy, Siena, Italy, [3]Genetica Medica Azienda Ospedaliera Univ.ria Senese Siena Italy, Siena, Italy, [4]Med. Genetics Univ. of Siena Italy; Med Biotech Hub and Competence Ctr. Dept. of Med. Biotechnologies Univ. of Siena Italy; Genetica Medica Azienda Ospedaliera Univ.ria Senese Siena Italy, Siena, Italy, [5]Unit of Ophthalmology and Retinoblastoma Referral Ctr. Dept. of Surgery Univ. of Siena Policlinico 'Santa Maria alle Scotte' Siena Italy, Siena, Italy, [6]Unit of Ophthalmology and Retinoblastoma Referral Ctr. Dept. of Surgery Univ. of Siena Policlinico 'Santa Maria alle Scotte' Siena Italy., Siena, Italy
Disclosure Block: **F. Privitera:** None.
Retinoblastoma (RB) is an ocular tumor of the pediatric age caused by biallelic inactivation of the RB1 gene (13q14). About 10% of cases are due to gross-sized molecular deletions. The deletions can involve the surrounding genes delineating a contiguous gene syndrome characterized by RB, developmental anomalies and peculiar facial dysmorphisms. Overlapping deletions previously found by traditional and/or molecular cytogenetic analysis allowed to define some critical regions for intellectual disability (ID) and multiple congenital anomalies, with key candidate genes. In the present study, we characterized by array-CGH seven new patients with interstitial 13q deletion involving RB1. Among these cases, three patients with medium or large 13q deletions did not present psychomotor delay. This allowed to define a minimal critical region for ID that excludes the previously suggested candidate genes (HTR2A, NUFIP1, PCDH8 and PCDH17). The region contains 36 genes including NBEA, which emerged as the candidate gene associated with developmental delay. In addition MAB21L1, DCLK1, EXOSC8 and SPART haploinsufficiency might contribute to the observed impaired neurodevelopmental phenotype. In conclusion, this study adds important novelties to the 13q deletion syndrome, although further studies are needed to better characterize the contribution of different genes and to understand how the haploinsufficiency of this region can determine ID.

PrgmNr 2609 - 7q36.1q36.2 deletion in a child with intellectual disability, developmental delay, short stature, and a Chiari 1 malformation

View session detail

Author Block: **E. Cordell**, Y-C. Hou, C. Gooch, J. Neidich; Washington Univ. Sch. of Med., St. Louis, MO
Disclosure Block: **E. Cordell:** None.
Introduction: Distal monosomy 7q36 is a rare chromosomal syndrome with symptoms including holoprosencephaly, growth restriction, developmental delay, intellectual disability, and facial features. In patients with distal monosomy 7q36, the partial deletion of the long arm of chromosome 7 includes the *SHH* gene located in 7q36.3 resulting in holoprosencephaly. Deletions in the proximal part of 7q36 are even less common. A case study from 2008 described a 9-year-old girl with a 5.27Mb deletion in 7q36.1q36.2 with a phenotype characterized by intellectual disability, facial features, renal hypoplasia, and long QT syndrome. Here we present a 12-year-old female with a 3.19MB *de novo* deletion in the 7q36.1q36.2 region. The patient presented with ADHD, autistic features, facial findings, intellectual disability, tremors, short stature, and a Chiari 1 malformation.
Methods: A chromosomal microarray analysis (CMA) was performed using the CytoScan HD array. Copy-number variant (CNV) interpretation guidelines were used for the evaluation and interpretation of CNVs. Results were confirmed using fluorescence *in situ* hybridization (FISH).
Results: The CMA test revealed an interstitial deletion of the long arm of chromosome 7 at band q36.1q36.2, arr[GRch37] 7q36.1q36.2(149828686-153017182)x1, approximately 3.18Mb in length. FISH analysis showed only one signal located at 7q36.1q36.2 confirming the terminal 7q deletion. The parental FISH were normal. This loss contained 45 protein-coding RefSeq genes and 7 disease-associated genes, *ASB10, CDK5, KCNH2, KMT2C, NOS3, PRKAG2* and *XRCC2*. Similar size losses have been documented as pathogenic or likely pathogenic in DECIPHER database in three individuals. Phenotypes included cardiac defects, congenital microcephaly, facial dysmorphisms, global developmental delay, and renal dysplasia. *KCNH2* and *KMT2C* genes have sufficient evidence for haploinsufficiency by ClinGen Dosage Sensitivity Map.
Discussion: The clinical features matched those of reported individuals with similar deletions encompassing the 7q36.1q36.2. She displays intellectual disability, short stature, and facial features consistent with the reported case. The patient has not developed long QT syndrome associated with loss of the *KCNH2* gene. Her recent ECG was normal. The patient has phenotypes associated with loss of the *KMT2C* gene, including autistic features, intellectual disability, short stature, and dysmorphic features. The patient has Chiari 1 malformation, not described previously. Our report contributes to the genotype-phenotype correlation in 7q36.1q36.2 deletion cases.

PrgmNr 2610 - Diagnostic yield of exome sequencing in neurodevelopmental disorders in a Brazilian clinical laboratory

View session detail

Author Block: **J. Sobreira**[1], D. Cupertino[1], D. Bermeo[1], T. Loureiro[1], D. Villela[1], G. Yamamoto[1], M. Gomes[1], R. Guarischi[1], C. de Souza[1], J. Silva[1], A. Bonaldi[1], A. Paes[1], G. Pereira[1], B. Schmabel[1], K. Endo[1], M. Milanezi[1], N. L. Sobreira[2], M. Migliavacca[1]; [1]LaboratÃ³rio GeneOne, SÃ£o Paulo, Brazil, [2]Johns Hopkins Univ, Baltimore, MD
Disclosure Block: **J. Sobreira:** None.
Neurodevelopmental disorders (NDDs) are a genetically heterogeneous set of conditions that affect 1-3% of children and is characterized by a wide range of phenotypic features, including intellectual disability (ID), autism spectrum disorder (ASD) and global developmental delay as the most common. Whole exome sequencing (WES), along with advanced bioinformatics capabilities, have created opportunities to use WES in a variety of medical situations, including for the molecular characterization of rare diseases such as neurodevelopmental disorders. A diagnostic rate between 25-38% has been reported for isolated NDD and between 41-64% for the syndromic NDDs. The diagnostic rate ranges from 29 to 50% among patients with ID, from 11 to 24% among patients with ASD and from 29 to 46% for patients with ID and ASD combined. Here we analyzed the diagnostic rate of WES performed by a Brazilian clinical laboratory in a cohort of 251 Brazilian patients with NDD. Among the 251 patients, 188 had the WES analyzed for the investigation of CNVs also. The patients were selected by searching the medical records for the following terms: 1) intellectual disability; 2) delay in global development, motor skills or speech, and; 3) hypotonia or muscle hypertonia or seizure (neonatal patients). The proportion between male and female was 143 (56.9%): 108 (43.1%), and the ages ranged from 29 days to 70 years. An overall diagnostic rate of 31.4% (79/251) was achieved, with the identification of two CNVs. The pathogenic variants were observed in 64 genes, with 59.7% related to autosomal dominant node of inheritance, 27.2% to autosomal recessive mode of inheritance and 13.1%. The diagnostic rate for CNV was 1%. Our data are consistent with previous reports and we observe that a better diagnostic rate is reported in cohorts where trio-WES sequencing is performed. The low diagnostic rate for large CNVs was already expected, since, in Brazil, investigation of patients with NDD using techniques such as array is required before performing WES.

PrgmNr 2611 - Genetic evaluation of children born with very low birth weight

View session detail

Author Block: **A. Jorge**[1], B. L. Freire[1], T. K. Homma[1], A. M. Lerario[2], I. J. P. Arnhold[1]; [1]Faculdade de Med. da Univ.e de SÃ£o Paulo (USP), Sao Paulo, Brazil, [2]Univ. of Michigan, Michigan, MI
Disclosure Block: **A. Jorge:** Consultant/Consulting Fees/Other Remuneration; NovoNordisk, Biomarin.
Very low birth weight (VLBW) children are defined as those infants with birth weight Patients and Methods: Forty-two children born VLBW referred for short stature investigation in a tertiary care hospital. Fourteen children were diagnosed based on clinical findings of a known condition and were submitted to a candidate gene approach: 6 patients were evaluated by a MS-MLPA due to clinical suspicion of Silver-Russel syndrome (SRS); and 8 cases were clinically diagnosed with a monogenic condition. Patients with a negative result in the candidate gene approach (n=5) or without clinical suspicion (n=28) were evaluated by chromosome microarray analysis (CMA; n=15) and/or whole exome sequencing (WES; n=25). Twelve patients had CMA negative results before performing WES. **Results:** All patients were born VLBW, 66% were VP, 74.5% born SGA and 72.5% has syndromic features. From the 42 patients who were selected for genetic analysis, a syndromic genetic condition was confirmed in 43% (18/42). A candidate gene approach confirmed the clinical diagnosis in 64.3% (9/14): SRS (n=4); pseudohypoparathyroidism (*GNAS*), resistance to IGF-1 (*IGF1R*), Bloom (*BLM*), Aarskog-Scott (*FGD1*) and Noonan (*SOS1*) syndromes (one each other). Twenty-eight patients without clinical suspicion and five patients with negative target analysis were selected to do genomic approaches. Among these patients, we diagnosed 27% (9/33) cases. Three patients had pathogenic CNVs and 6 patients had rare monogenic conditions: resistance to IGF1 (*IGF1R*), Microcephalic Osteodysplastic Primordial Dwarfism type II (*PCNT*), Centronuclear myopathy (*DNM2*), Baraitser-Winter (*ACTB*), Silver-Russell (*HMGA2*) and Hoyeraal-Hreidarsson (*TERT*) syndromes. **Conclusion:** A significant proportion of children born VLBW have a monogenic syndromic condition, with great heterogeneity of diagnoses including several rare conditions. The use of WES is an approach that allows the diagnosis of most cases, excluding patients with suspected epimutations defects.

PrgmNr 2612 - Lung Pathology in Takenouchi-Kosaki Syndrome: Expanding the phenotypic spectrum

View session detail

Author Block: **A. Ahmed**, R. Morgan, C. M. Oermann, I. Thiffault; Children's Mercy Hosp., Kansas City, MO
Disclosure Block: **A. Ahmed:** None.
Background: Takenouchi-Kosaki syndrome is a congenital malformation syndrome characterized by severe developmental delay, macrothrombocytopenia, camptodactyly, hearing loss and dysmorophic facial features associated with mutations in the *CDC42* gene. Lung pathology has not been previously described in the few cases that have been reported. **Case Report**: A female infant with past history of congenital heart disease, intestinal malrotation and developmental delay presented at 15 months of age with fever and respiratory distress. She was subsequently diagnosed with RSV bronchiolitis at 2 years of age. Work-up revealed chronic thrombocytopenia with giant platelets and immune deficiency with decreased numbers of T-lymphocytes and NK cells. Genomic sequencing from the patient and parents revealed the patient was heterozygous for a de novo pathogenic variant in *CDC42*, c.203G>A (p.Arg68Gln). She had frequent hospital admissions to intensive care for chronic respiratory distress, acute exacerbations and multiple episodes of viral infections. CT scan of the lung highlighted diffuse interstitial lung disease with prominent subpleural cystic changes. A wedge lung biopsy was processed for histopathology and ultrastructural morphology. There was diffuse interstitial fibrosis with patchy areas of chronic inflammation and emphysematous cystic changes. The dilated alveolar spaces displayed diffuse type II pneumocyte hyperplasia and increased alveolar macrophages. There was proteinaceous material associated with cholesterol clefts. Electron microscopy revealed many lamellar bodies, some of which showed electron dense deposits. Some lamellar bodies were abnormally small and dense and composite lamellar bodies were present. Although some of the lung pathology could be attributed to repeated infections, the presence of alveolar proteinosis, desquamative interstitial pneumonitis-like areas and ultrastructurally abnormal lamellar bodies suggested a surfactant deficiency disorder. The patient did not have any pathogenic variants in the known surfactant genes and other clinical and genetic features were more consistent with Takenouchi-Kosaki syndrome. **Conclusion**: This is the first case of the syndrome to be associated with severe chronic interstitial lung disease and morphologic surfactant abnormalities. The novel mutation site and the lung pathology further expand the genetic and phenotypic spectrum of this disease.

PrgmNr 2613 - Missense variants at a conserved 14-3-3 binding site in HDAC4 result in a novel intellectual disability syndrome

View session detail

Author Block: **E. Wakeling**[1], M. McEntagart[2], K. Stals[3], M. Bruccoleri[4], P. Campeau[5], A. Barnicoat[1], C. Beesley[1], T. Courtin[6], DDD study[14], S. Ellard[7], S. H. Elsea[8], A. Hanson-Kahn[9], T. Homfray[1], M. K. Kukolich[10], M. Morleo[11], C. Shaw-Smith[12], D. A. Stevenson[13], TUDP Study Group[14], M. Wakeling[14], X-J. Yang[4], R. Caswell[15]; [1]North East Thames Regional Genetic Service, London, United Kingdom, [2]St. George`s Univ. Hosp NHS Fndn. Trust, London, United Kingdom, [3]Exeter Genomics Lab., Exeter, United Kingdom, [4]Rosalind and Morris Goodman Cancer Res. Ctr. and Dept. of Med., Montreal, QC, Canada, [5]CHU Sainte-Justine Hosp., Montreal, QC, Canada, [6]Sorbonne UniversitÃ©, HÃ´pital de la PitiÃ©-SalpÃªtriÃ¨re, Paris, France, [7]Royal Devon & Exeter Hosp, Exeter, United Kingdom, [8]Baylor Coll. Med., Houston, TX, [9]Stanford Univ, Stanford, CA, [10]Cook Children's Hosp, Fort Worth, TX, [11]Telethon Inst. of Genetics and Med. (TIGEM), Naples, Italy, [12]Dept. of Clinical Genetics, Royal Devon and Exeter NHS Fndn. Trust, Exeter, United Kingdom, [13]Univ of Utah, Salt Lake City, UT, [1415]Inst. of BioMed. and Clinical Sci., Univ. of Exeter Med. Sch., Exeter, United Kingdom

Disclosure Block: **E. Wakeling:** None.

Histone deacetylases (HDACs) have a central role in the regulation of chromatin structure and gene expression in the eukaryotic cell. Disruption of their activity has been shown to cause a wide range of developmental disorders. Loss-of-function alleles of *HDAC4*, a class IIa deacetylase, have been reported in Brachydactyly-mental retardation syndrome (BDMR). Loss of HDAC4 function usually occurs as part of longer deletions of chromosome 2q37 and BDMR is also known as Chromosome 2q37 deletion syndrome. The precise role of HDAC4 within this phenotype remains uncertain. Eight unrelated patients with a novel phenotype distinct from BDMR and heterozygous *de novo* missense variants in *HDAC4* gene were identified. Common features in these patients include severe intellectual disability, epilepsy, swallowing difficulties, scoliosis and distinctive facial features. All eight have variants within a key regulatory site spanning residues 242-248 of HDAC4. This site binds to 14-3-3 proteins and regulates nucleocytoplasmic shuttling of HDAC4. Six individuals carried variants affecting Pro248, altering a key residue for 14-3-3 binding. Two patients carried variants p.(Thr244Lys) and p.(Glu247Gly) which were shown by immunoprecipitation assay to impair 14-3-3 binding. We propose that reduced 14-3-3 binding results in enhanced nuclear import, in turn leading to a novel gain-of-function effect. Two further individuals were identified with variants in the C-terminal region of *HDAC4*, in or immediately before the nuclear export signal (p.(Glu1049Lys), p.(Thr1055Lys)). Both patients have a similar phenotype, including severe developmental delay and intellectual disability. Our hypothesis is that these two variants give rise to a similar increase in nuclear HDAC4, as a result of impaired nuclear export. Overall these findings suggest that increased nuclear activity of HDAC4 is the underlying cause of a newly characterised developmental disorder.

PrgmNr 2614 - Molecular and clinical diagnosis of 17p11.2 microduplication described as Potocki-Lupski Syndrome: Case Report

View session detail

Author Block: **M. Mesa**[1], D. NuÃ±ez[2], C. Lattig[3], A. Lopez[4]; [1]Facultad de Med., Univ. de los Andes, BogotÃi D.C, Colombia, [2]Univ. de los Andes, BogotÃi D.C, Colombia, [3]Univ. de los Andes, Bogota, Colombia, [4]FundaciÃ³n Santa Fe de BogotÃi, Bogota, Colombia
Disclosure Block: **M. Mesa:** None.
Background: Potocki-Lupski Syndrome (OMIM #610883) is a rare genetic disorder that affects 1 in 25,000 people and is genetically associated with a heterozygous duplication at chromosome 17p11.2 that encompasses *RAI1.* This syndrome is autosomal dominant, but it has been reported that most cases arise from de novo mutations. Phenotypically, it presents as a developmental disorder characterized by congenital anomalies along with neurodevelopmental delays such as hypotonia, failure to thrive, mental retardation and pervasive developmental disorders, accompanied with growth hormone deficiency.
Case Presentation: We report the clinical and molecular characterization of a 13 year old girl with 17p11.2 microduplication associated with atypical autism spectrum disorder, mental retardation and renal agenesia. Initially, the patient was tested for Angelman-gene *UBE3A* (FISH) with a negative result due to a phenotype characterized by unmotivated laughs, stereotyped movements, and sleep disorder. The patient was evaluated by the CGH + SNP array which revealed chromosomic alteration consistent with a mosaic duplication of 3.6mpb in chromosome 17p11.2 that encompasses 61 genes compatible with Potocki-Lupski syndrome. It was until one year after her Potocki-Lupski Syndrome diagnosis, that her parents reported symptoms of urinary incontinence throughout her entire life and along with diagnostic images she was diagnosed with renal agenesia, so she was referred to other specialists to keep the follow up of her disease.
Discussion: Here we present a patient diagnosed with Potocki-Lupski Syndrome where molecular testing played a big role in the diagnosis and the further follow up of the patient. The patient was referred to other specialists along with clinical genetics to follow up her condition and give her the best type of care. Although, our patient was diagnosed at 13 years-old, it is important to start testing for chromosomic alterations along with other complementary tests in patients that present phenotypic anomalies as soon as possible to give them an interdisciplinary treatment in order to improve their symptoms and their quality of life.

PrgmNr 2615 - Automated variant classification workflows of known and novel variants can be used to maintain quality standards, support standardization and reduce turn-around times in a rare disease laboratory

View session detail

Author Block: **H. Savage**; Congenica Ltd, Hinxton, United Kingdom
Disclosure Block: **H. Savage:** Salary/Employment; Congenica Ltd.
As demand for genomic testing increases, driven by greater accessibility and falling sequencing costs, variant analysts are encountering increasingly complex patients in addition to their routine cases. While increased access to testing is to be celebrated, the use of exome and genome testing for a wider range of disorders means variant interpretation is a key bottleneck, as even highly skilled and experienced analysts are not experts on all genes/disorders encountered. Yet this is the daily interpretation challenge we face.
Achieving a timely diagnosis has a huge impact on patients and their families; from providing answers to questions such as â□□why am I or my child different?â□□, to informing reproductive choice, providing access to support for patients, information about prognosis and early treatment leading to improved patient outcomes. However, the increased amount of time needed to familiarize analysts with novel genes and variants can increase turnaround times and may lead to delays in reporting results. With a finite workforce laboratories may face a situation where lack of resources for analysis of complex cases could lead to an extended diagnostic odyssey for patients and their families. This issue was recently confirmed in a survey of Clinical Laboratories which identified that 71% are at, or approaching, capacity.
Automated processes are commonplace in the diagnostic laboratory; from liquid-handling robotics to automated bioinformatics pipelines processing large volumes of data, however automated interpretation is yet to be widely adopted. As clinical variant analysis can take up 11 hours in just simple rare disease cases, with complex cases taking up to 16 hours, unsupported manual interpretation is not a scalable solution. Is it time to embrace automation of pre-classified variants, automated assignment of ACMG criteria to novel variants, and integration of powerful HPO-driven ranking algorithms to support timely diagnosis for even the most complex of patients?
Here I present the case for automating standardized analysis of cases to rapidly generate high quality, relevant results to support the diagnosis of patients with rare genetic disease, without compromising the diagnostic yield or quality of each analysis.

PrgmNr 2616 - Brazil enters the genome sequencing era: The Rare Genomes Project - interim results

View session detail

Author Block: R. M. Minillo[1], E. Perrone[1], C. R. D. Quaio[1], L. A. Virmond[1], K. Chen[2,1], C. A. Moreno[1], J. M. Ceroni[1], T. F. Almeida[1], A. V. C. Coelho[1], N. B. Zurro[1], G. B. C. Colichio[1], R. S. Reis[1], M. C. Cervato[1], M. P. Caraciolo[1], R. B. Alexandre[1], F. C. Val[1], G. S. Oliveira[1], D. D. C. Oliveira[1], L. M. S. Moura[1], P. N. Chazanas[1], M. Gretschischkin[1], B. M. C. Azevedo[1], M. S. Nobrega[1], C. L. P. Mangueira[1], **J. O. Filho**[1]; [1]Albert Einstein Hosp., SÃ£o Paulo, Brazil, [2]SÃ£o Paulo, Brazil

Disclosure Block: **J.O. Filho:** None.

Rare diseases (RD) are estimated to affect between 3.2 and 13.2 million individuals in Brazil. The Rare Genomes Project (RGP) is a Brazilian initiative for whole genome sequencing (WGS) of patients with RD recruited from centers for rare diseases of the National Unified Health System (SUS). RGP initiated in 2020 and will sequence the genomes of 7755 patients with RD until the end of 2023, allowing the creation of the largest Brazilian genomic database of patients with RD and hereditary cancer. RGP also aims to study the clinical and social burden of RD in Brazil, provide valuable information for cost-effective national health policies for a population generally underrepresented in previous genomic studies and improve diagnosis, therapy, prevention, genetic counseling and the quality of life of patients. The ongoing project has recruited 1486 probands from nine centers of four out of five macroregions of the country. Among those, approximately 1300 have already had their genomes sequenced, and for 163 clinical reports were returned to attending physicians.The most prevalent disorders are neurological (n=402, average age:11.8 years), congenital malformations (n=376, average age:10.6 years), hereditary cancer (n=343; average age:45), inborn errors of immunity (n=145; average age:19.6 years) and clinical genetic syndromes (n=114; average age:10.9 years). Out of the 163 reports released, molecular diagnosis was established for 57 patients (34,9%). In 43 patients, no candidate variants were detected (26,38%). The remaining 63 (38,7%)patients presented variants of unknown significance (VUS) in heterozygous or homozygous state, heterozygous state for a recessive phenotype or pathogenic/likely pathogenic (P/LP) variants in heterozygous state and a VUS in the same gene for a recessive phenotype. Among the positive cases, the higher diagnostic rates were seen in the neurological (18 cases; 36%), clinical genetic syndromes (10 cases; 17.6%), inborn errors of metabolism (7; 12.3%), immunodeficiencies (5; 8.8%), and hereditary cancer syndromes (5;8.8%) cohorts. A total of 81 variants were detected in the 57 positive patients (67 SNVs and 14 large-scale CNVs/SVs), 69 (85.2%) being annotated as (P/LP) (57 SNVs and 12 CNVs/SVs). Interesting to mention that 112 positive cases (21.1%) had P/LP CNVs or SNVs in regulatory regions which would not be identified by whole exome sequencing (WES). In summary, the RGP is the largest rare disease sequencing effort in Brazil, and the preliminary results demonstrate a high diagnostic rate.

PrgmNr 2617 - Clinical relevance of the study of pigmentary mosaicism using genomic *array*

View session detail

Author Block: **Y. G. Gasparini**[1], M. M. Montenegro[1], A. M. Nascimento[1], V. T. Almeida[1], G. F. S. Carvalho[1], S. N. Chehimi[1], B. M. Wolff[1], M. S. Caravante[1], C. A. Kim[2], L. D. Kulikowski[1]; [1]Laboratorio de Citogenomica, Departamento de Patologia, Faculdade de Med. FMUSP, Univ.e de Sao Paulo, Sao Paulo, Brazil, [2]Inst da Crianca, Sao Paulo, Brazil
Disclosure Block: **Y.G. Gasparini:** None.
The mosaicism is a usual event in the world population and is defined by the presence of two or more genetically different cell lines in the same individual, providing for the same zygote. However, the identification and characterization of mosaicism is the big challenge of the area laboratorial medicine, because maybe the tissue canâ□□t reveal the variant in mosaicism. This is associated with mosaicism confined to tissue or pigmentary mosaicism, that constitutes a heterogeneous group of skin pigmentation alterations associated with multisystem involvement. Thus, the nuances of these clinical observations were prescient and resulted in the construction of several hypotheses based on the idea that the concomitant genomic alterations in different tissues could explain unusual patterns of phenotype within the same disease. We report on an eleven-year-old boy with mild intellectual disability, short neck, congenital cervical block, hyperchromic spots in Blaschko lines, JUP stenosis on the left, intestinal constipation, clinodactyly of the hands and enlarged halluces. We have collected samples from peripheral blood, saliva and skin fibroblasts. Results from cytogenetic analysis (GTG Band) performed in fibroblasts obtained from spots and peripheral blood were normal. Unexpected results from skin fibroblasts, peripheral blood and saliva using genomic array (cytoSNP850k-Illumina) revealed a pathogenic duplication in 17q21.31 in mosaic. The region that involving the duplication in 17q21.31 contemplate the two important genes, KANSL1 and KANSL1-AS1, that associated a Koolen-De-Vries syndrome (#610443) and 17q21.31 Duplication Syndrome (#613533) and in both is related the clinical phenotype similar to that presented by the patient is reported like as constipation intestinal, JUP stenosis, clinodactyly and mild intellectual disability. Our results highlight the importance of genomic arrays to investigate patients with pigmentary mosaicism to assess genomic variation among, and increasingly within, individuals and improve the genotype-phenotype correlations.

PrgmNr 2618 - Fold-back mechanism originating an inv dup del rearrangement in chromosome 15

View session detail

Author Block: **B. Burssed**, F. T. Bellucco, M. Zamariolli, B. P. Favilla, M. I. Melaragno; Div. of Genetics, Dept. of Morphology and Genetics, Univ.e Federal de SÃ£o Paulo, SÃ£o Paulo, Brazil
Disclosure Block: **B. Burssed:** None.
Inverted duplications associated with terminal deletions characterize a family of rearrangements known as inv dup del. Four mechanisms have been proposed to explain their formation and three of them lead to the appearance of a normal copy number region (spacer) between the two copies of the duplicated segment. In this study, we investigated a patient with facial dysmorphisms, underdeveloped nasal alae, micrognathia, hypothyroidism, hearing loss, hydrocephalus, and brachydactyly with an intra-arm rearrangement in chromosome 15q. Karyotyping and CytoScan HD Array (Affymetrix) revealed a de novo 19.5 Mb duplication and a 5.5 Mb deletion in 15q, as follows: arr[GRCh37] 15q24.3q26.2(77348331_96973582)Ã□3,15q26.2q26.3(96973928_102429040)Ã□1. Custom array (Agilent), with enrichment of probes from chromosome 15, was performed to determine the putative breakpoint with more precision and presented the following result: arr[GRCh37] 15q24.3q26.2(77338889_96963865)Ã□3,15q26.2q26.3(96966519_102263635)Ã□1. A normal copy number region was found between the duplication and the deletion for both catalog and custom arrays, with sizes of 346 bp and 2,654 bp, respectively. Long-range PCR revealed an inversion of the duplicated segment, and Sanger sequencing confirmed that the patient has an inv dup del(15q) with the presence of a 2,743-bp spacer. The breakpoint between the end of the spacer (chr15:96966212) and the start of the inverted duplication (chr15:96963469), which were much closer to the ones shown by the custom array than to the catalog array, presented four nucleotides of microhomology. Based on these results, we proposed that the rearrangement was formed through the Fold-back mechanism, in which a double-strand break leads to a 5' resection, which then exposes microhomologies in the 3' free end, allowing the chromosome to fold onto itself. This fold-back loop region that facilitates the intrastrand pairing within a sister chromatid corresponds to the spacer. DNA replication forms an unstable dicentric, which will later be broken to form the inv dup del chromosome. The precise determination of the genomic regions subjected to copy number variation enabled a better understanding of the patient's phenotype. Financial support: SÃ£o Paulo Research Foundation (FAPESP), Brazil.

PrgmNr 2620 - Incorporating a risk modifying SNP (rs143838139) to increase the detection of [2+0] carriers for Spinal Muscular Atrophy

View session detail

Author Block: **G. Ware**[1], C. Miller[1], D. Jones[1], m. avenarius[2]; [1]The Ohio State Univ., Columbus, OH, [2]The Ohio State Univ., Columbus, OH
Disclosure Block: **G. Ware:** None.
Spinal muscular atrophy (SMA) is an autosomal recessive motor neuron disease that most commonly presents in infancy with hypotonia. SMA is caused by biallelic inactivation of the survival motor neuron 1 (*SMN1*) gene, with most patients showing homozygous deletions and up to 10% showing one intragenic mutation and one deletion. The severity of this disease is inversely modified by increasing copies of *SMN2*, a hypofunctional paralog of the *SMN1* gene. Recently, SMN-enhancing therapies have become available and are linked to remarkable improvements in clinical presentation. With a prevalence of ~1 in 10,000 live births (carrier frequency of ~1:50), SMA is one of the most common severe childhood-onset diseases; therefore, current guidelines recommend SMA carrier screening before or during pregnancy. At our institution, we employ a competitive quantitative PCR method, including *SMN1* and *SMN2* standards and a co-amplified control gene. This gene dosage assay detects ~96% of all SMA carriers but not the remaining 4% of carriers who harbor two copies of *SMN1* arrayed in -cis. Although definitive detection of these [2+0] carriers is challenging, it has been noted that the c.*3+80T>G (rs143838139) variant is positively correlated with this chromosomal configuration the degree to which depends upon racial/ethnic background. Bayesian analysis is used to calculate the residual risk of being a carrier with counseling and paternal testing offered to those couples with higher risk. In July 2020, we enhanced our standard SMA carrier testing by adding detection of this risk-modifying SNP. Detection of this SNP relies on *SMN1*-specific amplification of intron 7, restriction digestion with HpyCh4III, and fragment sizing by capillary electrophoresis. Since implementing this enhanced assay, we have screened 1,117 patients. 21 (1.9%) patients tested positive for a single copy of *SMN1* and 163 showed the rs143838139 variant (14.6%). Among the latter group, 36 (22%) harbored two copies of *SMN1* and were reported as potential [2+0] carriers for SMA. The remaining risk-modifier positive patients (n=127) showed a range of *SMN1* copy numbers including 1 patient with 1 copy of *SMN1*, 98 patients with 3 copies of *SMN1*, and 28 patients with 4 copies of *SMN1*. At this time, the risk-modifying SNP is not informative for these individuals and was not reported for these patients. As allele frequency data is limited for this SNP, our study may contribute additional population-level genetic data to help increase the detection of carriers for SMA in other patient populations. Additionally, studies are ongoing to understand how our enhanced SMA carrier screen is utilized in clinical practice at our institution.

PrgmNr 2622 - The ACMG SF v3.0 gene list increases actionable variant detection by 25% over v2.0 in the ClinSeqÂ® cohort

View session detail

Author Block: **J. J. Johnston**, M-L. Brennan, K. B. Radenbaugh, S. Yoo, S. Hernandez, K. L. Lewis, A. Katz, T. A. Manolio, L. G. Biesecker; NHGRI, Bethesda, MD
Disclosure Block: **J.J. Johnston:** None.
Background: The American College of Medical Genetics and Genomics (ACMG) recommends interrogation of a minimum gene list for return of secondary findings from clinical exome and genome sequencing. The ACMG has recently released the third version of this list (ACMG SF v3.0) with the addition of 14 genes that meet the ACMG Secondary Findings Working Groupâ□□s requirements for actionability, severity, penetrance, and impact or burden of treatment and/or screening. These genes, known to be causative of disorders with defined phenotypes that are clinically actionable, are associated with cardiovascular (*TTN*, *FLNC*, *TRND*, *CASQ2*), cancer susceptibility (*PALB2*, *MAX*, *TMEM127*), inborn errors of metabolism (*GAA*, *BTD*), and other (*ACVRL1*, *ENG*, *RPE65*, *HFE*, *HNF1A*) phenotypes. **Methods:** To determine the additional variant yield in a cohort of unselected individuals we interrogated the ClinSeqÂ® cohort (~1450 individuals, 57% White, 32% African American or Black, 10% other) for variants in these 14 genes. We restricted our analyses to coding variants, +1,+2/-1,-2 splice site variants, and the pathogenic *GAA* variant, NM_000152.5:c.-32-13. Variants were assessed according to ACMG/AMP guidelines, including manual curation of 305 unique variants for evidence of pathogenicity. Forty-three individuals had potential loss of function variants in *TTN* (33 unique variants) which far exceeded the expected frequency based on the prevalence of dilated cardiomyopathy. Nineteen *TTN* variants in exons included in less than 80% of transcripts were excluded from the analysis. Another 11 *TTN* variants were excluded due to low coverage resulting in the variant being represented by a single paired-end read. **Results:** A total of 27 unique variants (in 185 individuals) were classified as pathogenic/likely pathogenic (P/LP). Nine individuals had P/LP variants identified in the heterozygous state in genes associated with disease with autosomal dominant inheritance, *FLNC* (1), *HNF1A* (1), *PALB2* (3), *TMEM127* (1), *TTN* (3). *HFE*, p.Cys282Tyr was identified in the homozygous state in four individuals. Seventeen P/LP variants (in 172 individuals) were identified in the heterozygous state in genes associated with disease inherited in an autosomal recessive pattern and were not considered for return. Excluding the four individuals with *HFE*, p.Cys282Tyr, which is known to be more common in individuals with European ancestry, the frequency of returnable P/LP variants did not differ by ancestry. **Conclusion:** Using the updated ACMG v3.0 gene list, the number of individuals with a returnable P/LP variant increased in the ClinSeqÂ® cohort by 25%, from 3.7% (n=53, ACMG SF v2.0) to 4.5% (n=65, ACMG SF v3.0).

PrgmNr 2623 - The knowns and unknowns: Challenges of Genes of Uncertain Significance in the era of high-throughput sequencing and lessons from Matching Tools

View session detail

Author Block: **I. Thiffault**[1], E. G. Farrow[2], C. Saunders[2], A. S. Cohen[3], J. Alaimo[1], F. Del Viso[1], V. Paolillo[1], T. Zion[1], S. Neal[1], T. Pastinen[3]; [1]Children's Mercy Hosp., Kansas City, MO, [2]Children's Mercy Hosp, Kansas City, MO, [3]Children's Mercy Kansas City, Kansas City, MO
Disclosure Block: **I. Thiffault:** None.
Gene-disease associations continue at a high rate in the era NGS testing. However, many genes lack sufficient evidence to be considered clinically-relevant. A scoring system for gene-disease associations was proposed by ClinGen, considering experimental/genetic evidence, such as the number of probands with compelling variants. Extremely rare variants, including de novo, in candidate genes may be implicated in disease, but are reported as â□□genes of unknown significanceâ□□ (GUS) in our center when evidence is lacking. In the context of a health system wide genomic medicine program Genomic Answers for Kids (GA4K) and our clinical NGS testing program at Childrenâ□□s Mercy, we performed a retrospective study of variants detected in GUS of all patients/families (>5000) tested for suspected genetic disorder. Since 2015, we submitted >550 GUS in GeneMatcher. The rate of GUS detection has significantly increased with ~ 3 GUS per week and an annual Matching tools-related publication rate of 10 manuscripts a year. We submitted >1 GUS in a subset of proband [â□¥2 (11%); â□¥3 (2%)] for which the variants were de novo and/or more than one genetic etiology were suspected. The variants included in this study are all absent or nearly absent from databases. More than 75% of the GUS were associated with de novo variant and only 2% were recurrent in unrelated families within our cohort â□□highlighting the need for data sharing. We experienced a high rate of â□□Matching hitsâ□□ (Nâ□¥10) only for GUS with de novo and/or loss-of-function variants. There is a positive correlation with the number of matching entries and the rate of publication. X-linked (XL) or autosomal recessive (AR) GUS (~25% of our entries) are underrepresented in Matching tools with an ~ 0-3 hits. Moreover, XL and AR hits had ~ 2.5 years length for functional characterization and publication. A subset of GUS (5%) are OMIM genes with novel phenotype/inheritance. The vast majority of our cohort were referred for neurological disorder with or without congenital anomalies (~75%). Many of the genes have only been reported in large autism cohorts where limited phenotypic information is available. The sharing of carefully vetted GUS, as in this study, is highly important to establish solid gene-disease relationships and molecular diagnosis. Thus, additional cases with phenotypic data are imperative to upgrade GUSs to disease causing status but also to expand phenotypic spectrum of known-OMIM genes. We propose that speed and yield of new gene discovery can be enhanced by additional methods of data sharing, which we are currently implementing via a registered access phenotype/genotype GA4K rare disease resource on a cloud-based PhenoTips platform.

PrgmNr 2624 - What does a 45,X karyotype add to 5p- syndrome literature?

View session detail

Author Block: **S. Chehimi**[1], V. Almeida[1], A. M. Nascimento[1], Y. G. Gasparini[2], G. S. Carvalho[2], B. M. Wolff[1], M. M. Montenegro[1], N. A. Assuncao[3], C. A. Kim[4], L. D. Kulikowski[5]; [1]Faculdade de Med. da Univ.e de SÃ£o Paulo, Sao Paulo, Brazil, [2]Faculdade de Med. da Univ.e de SÃ£o Paulo, SÃ£o Paulo, Brazil, [3]Escola Paulista de Med., Univ.e Federal de Saì□o Paulo, Sao Paulo, Brazil, [4]Inst da Crianca, Sao Paulo, Brazil, [5]Univ. de Sao Paulo, Sao Paulo, Brazil
Disclosure Block: **S. Chehimi:** None.
Rearranged chromosomes may often lead to distress in a genetic diagnosis. Some specific monosomies, such as in 5p- syndrome, with a well-described phenotype, may undergo undiagnosed when there are other autosomes or sexual chromosomes involved. Here we describe a rearrangement between chromosomes 5 and Y that could only be seen with the combination of different molecular techniques that characterized the chromosomes rearrangements. The patient investigated was 5-years old and presented with a normal male genitalia. The first molecular approach was the karyotype that indicated the result of 45,X. MLPA and FISH indicated that there were deletions in 5p and Yq probes with a positive SRY signal in 5p region (FISH result was 45,X.ish X(DXZ1+),der(5)(SRY+)). Array was a complementary molecular test that confirmed previous results and defined the size of the deletions (arr[GRCh37] 5p15.33p15.1(22149_17812007)x1; Yq11.21q12(14852740_59335913)x0). Gathering all the results, we have inferred that one of the pair of chromosomes 5 was derived from 5p- chromosome and Yq- chromosome and this rearrangement probably generated a stable dicentric derivative chromosome. We could only find two reports on patients with t(Y;5) and both were published more than 30 years ago (Vignetti et al., 1977; Weber et al., 1987) with no molecular characterization. Generally, microdeletions on the Y chromosome are associated with male infertility and azoospermia, but previous studies involving translocations between chromosomes 5 and Y do not point to any specific changes associated with the Y deletion (including reports mentioning normal external male genitalia and no other abnormalities). Previous studies also pointed to phenotypic characteristics commonly associated with 5p-, such as hypertelorism and micrognathia. In conclusion, we have observed that 5p- syndrome was already described in other rearrangements with different autosomes and chromosomes but this case is unique, so far, indicating the need for a complete molecular characterization to follow a proper genetic familiar counseling.

PrgmNr 2625 - *APOE4* increases astrocyte reactivity in cells with local European ancestry

View session detail

Author Block: **O. Oron**[1], J. P. Rodgers[1], Y. Akvan[1], M. L. Vasquez[1], K. Celis[1], K. S. Maddy[1], C. M. Chambers[1], B. E. Feliciano-Astacio[2], G. W. Beecham[1], M. L. Cuccaro[1], H. N. Cukier[1], M. Blurton-Jones[3], M. A. Pericak-Vance[1], J. M. Vance[1], D. M. Dykxhoorn[1]; [1]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Univ. Central del Caribe, BayamÃ³n, PR, [3]Dept. of Neurobiology and Behavior, Univ. of California, Irvine, Irvine, CA

Disclosure Block: **O. Oron:** None.

Recently we demonstrated that local genomic ancestry (LA) drives the difference in AD risk between European (EU) and African (AF) carriers of *APOEÎµ4/4*. As a follow-up study using single-nuclei RNAseq, we found that AD *APOEÎµ4* homozygotes with EU Local Ancestry (LA) had a significantly increased *APOEÎµ4* expression compared to AD *APOEÎµ4* homozygotes with AF LA. In two of the EU LA brains, an astrocyte cluster which showed the highest *APOEÎµ4* expression was observed, also expressed a panel of genes consistent with A1 Reactive Astrocytes (A1RA). No such cluster was seen in the AF LA brains. A previous study in mice suggested *APOEÎµ4* maybe a contributor to A1RA development. We sought to explore the relationship of increased *APOEÎµ4* expression and A1RA using inducible Pluripotent Stem Cells (iPSC)-derived astrocytes.

iPSC lines were reprogrammed from PBMCs derived from one EU LA patient, and one AF LA patient, bearing the *APOEÎµ4/4* genotype. They were subsequently validated for pluripotency by Immunocytochemistry (ICC) of NANOG, SOX2 and OCT4, and genomic stability. Following differentiation into astrocytes, they were validated by GFAP and Î²S100 ICC. Upon maturity (Day 54 in-vitro), astrocytes were treated with either a cytokine cocktail (IL-1a, hTNFa, C1q) or overexpressed with *APOEÎµ4* by lentiviral transduction for 14 days (two replicates per line). mRNA was extracted and qPCR was performed to measure changes in *APOEÎµ4* and markers of A1RA (C3, GBP2, IFITM3). Astrocytes treated with the cytokine cocktail had a 100- and 600-fold increase in C3 in EU and AF LA astrocytes, respectively. Upon transgenic *APOEÎµ4* overexpression (OE), a significant increase in C3 was observed in the EU LA astrocytes, while no increase was observed in the AF LA astrocytes.

Our preliminary results support the hypothesis that APOEÎµ4 OE can increase the conversion of astrocytes to the toxic A1RA state in EU LA astrocytes. Additional EU LA and AF LA astrocytes are currently being investigated to further investigate this initial finding.

PrgmNr 2626 - A murine model lacking *Lyst* displays progressive neurodegeneration and recapitulates Chediak-Higashi Syndrome

View session detail

Author Block: **S. Greene**[1], J. D. Burke[1], E-R. Nicoli[2], M. Morimoto[2], P. M. Zerfas[3], L. Garrett[4], G. Elliot[4], C. Toro[5], W. A. Gahl[6], W. J. Introne[7], M. Malicdan[5]; [1]Human Biochemical Genetics Section, Med. Genetics Branch, NHGRI, NIH, Bethesda, MD, [2]Undiagnosed Diseases Program, Common Fund/ NHGRI, NIH, Bethesda, MD, [3]Diagnostic & Res. Services Branch, OD, NIH, Bethesda, MD, [4]Transgenic Mouse Core, NHGRI, NIH, Bethesda, MD, [5]NIH, Bethesda, MD, [6]NHGRI (NIH), Bethesda, MD, [7]NIH, Bethesda, MD

Disclosure Block: **S. Greene:** None.

Background: Chediak-Higashi Syndrome (CHS) is a rare autosomal recessive disease with congenital immunodeficiency, bleeding diathesis, partial oculocutaneous albinism, and progressive neurologic impairments that include parkinsonism, ataxia, and sensorimotor neuropathies. CHS is due to bi-allelic mutations in the lysosomal trafficking regulator gene (*LYST*), which encodes a 429-kDa protein. Patient-derived cells are noted to have enlarged lysosomes clustered in the perinuclear region, indicating LYST might be involved in lysosomal fusion or fission. Despite *LYST* mutations being recognized as the cause of CHS more than a decade ago, a knowledge gap remains on the exact function of LYST and how mutations in *LYST* cause CHS. This knowledge gap is likely due to several factors, including the absence of an animal model that faithfully recapitulates all aspects of CHS. **Methods:** We generated a knockout mouse model ($Lyst^{-/-}$) that lacks a functional murine *Lyst* homologue. Using CRISPR/Cas9 technology, we deleted exons 4-53, creating a *Lyst* knockout mouse model, $Lyst^{-/-}$. Phenotypic characterization was performed to establish the molecular, biochemical, and histological consequences of deleting the majority of the *Lyst* gene. We compared our results to the phenotype of the existing $Lyst^{bg}$/J, a spontaneous murine mutant that has a homozygous 3 bp deletion in the penultimate exon of the gene but lacks a robust neurologic phenotype. Extensive murine behavioral tests were employed, focusing on the neurological impairments in mice. **Results:** $Lyst^{-/-}$ mice are viable, present with lighter pigmentation of the coat and skin of the ears and tail, and exhibit prolonged bleeding times. Measurement of mRNA expression in multiple tissues revealed markedly reduced *Lyst* levels. Mouse embryonic fibroblasts from $Lyst^{-/-}$ mice depict enlarged, perinuclear lysosomes and polymorphonuclear leukocytes contain enlarged granules. $Lyst^{-/-}$ mice develop ataxia and declining motor coordination with signs of an intention tremor by 9 months of age, compared to the late onset of ataxia in $Lyst^{bg}$/J mice (17 months). Histologically, $Lyst^{-/-}$ mice show significant Purkinje cell loss at 12 months of age, and evidence of a progressive peripheral neuropathy from 3 months to 24 months of age. **Conclusions:** Our $Lyst^{-/-}$ model replicates many important aspects of the human CHS phenotype, most notably neurological manifestations that parallel the human disease. Our $Lyst^{-/-}$ mice represent a system on which to study the basis of neurodegeneration and explore the potential therapeutic efforts for this devastating disease.

PrgmNr 2627 - A splice site SNP in *Gabra2* is a modifier of *Scn8a* encephalopathy in the mouse

View session detail

Author Block: **W. Yu**[1], M. K. Mulligan[2], R. W. Williams[2], M. H. Meisler[1]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN
Disclosure Block: **W. Yu:** None.
De novo gain-of-function mutations of *SCN8A* are responsible for early onset developmental epileptic encephalopathy (OMIM #614558). Severely affected individuals exhibit refractory seizures, developmental delay, and cognitive disabilities, often accompanied by impaired movement. Patients carrying the same variant can differ in clinical course, suggesting a role for modifier genes in disease severity. We used a mouse model of the patient *SCN8A* mutation p.Arg1872Trp in the C57BL/6J strain background to screen for genetic modifiers by carrying out crosses with other inbred strains. In an F2 cross with strain SJL/J, we observed mice with increased age of seizure onset and prolonged survival. QTL analysis identified a modifier that mapped to an 15 Mb region of mouse chromosome 5 containing the *Gabra2* gene (Yu et al, *Epilepsia* 61:2847-2856, 2020). C57BL/6J mice carry a hypomorphic mutation at the exon 5 splice acceptor site of *Gabra2* that results in a 4x reduction of transcript abundance (Mulligan et al, *Front. Genet.* 10:188, 2019). Homozygosity for the hypomorphic C57BL/6J allele was associated with early seizure onset and shortened life span in the F2 mice. We have now confirmed *Gabra2* as the modifier gene by analysis of a cross with a corrected line of C57BL/6J in which the deleted T nucleotide at position -5 upstream of exon 5 has been replaced (Mulligan et al, *Front. Genet.* 10:188, 2019). Correction of the deletion restores transcript abundance and also restores the later age of seizure onset. *Gabra2* encodes the alpha2 subunit of the GABA receptor that provides inhibitory input to the axon initial segment of excitatory neurons. Reduced inhibitory input due to the hypomorphic C57BL/6J allele is consistent with elevated neuronal excitability and the enhanced seizure phenotype in homozygotes. The 'private' mutation of *Gabra2* in strain C57BL/6J could influence other studies of neurological phenotypes. There is a potential role for quantitative variation in human GABA receptor expression on the severity of genetic epilepsies as well as therapeutic interventions.

PrgmNr 2628 - Characterization of the genetic architecture of neurodevelopmental disorders using Duchenne Muscular Dystrophy as a model

View session detail

Author Block: **C. Samogy Costa**[1,2], L. Madanelo[2], E. Zachi[2], G. Yamamoto[2], A. Girardi[2], J. Y. T. Wang[2], E. Vadasz[3], G. Campos[1], A. C. Krepischi[1], M. Zatz[1,2], M. R. Passos-Bueno[1,2]; [1]Inst. de BiociÃªncias, Univ.e de SÃ£o Paulo, SÃ£o Paulo, Brazil, [2]Centro de Estudos do Genoma Humano e CÃ©lulas-Tronco, Univ.e de SÃ£o Paulo, SÃ£o Paulo, Brazil, [3]Inst. de Psiquiatria - Hosp. das Clinicas da Faculdade de Med., Univ.e de SÃ£o Paulo, SÃ£o Paulo, Brazil
Disclosure Block: **C. Samogy Costa:** None.
Duchenne muscular dystrophy (DMD) is a muscular disease caused by loss of function variants in the *DMD* gene. In addition to the muscle phenotype, neurodevelopmental disorders, such as intellectual disability (ID) and autism spectrum disorders (ASD) are more frequent among DMD individuals than in the general population. In this context, different studies have shown the relevance of rare variants of moderate effect in the genetic background for the manifestation and variability of NDDs, compatible with an oligogenic inheritance model. Our work aims to investigate whether mutations in *DMD*, associated with an accumulation of rare variants with moderate effect on neurodevelopmental genes, are responsible for the development of ASD and/or ID in these individuals. To achieve this goal, we are setting up a group of 57 DMD individuals for which ASD and ID were evaluated. Clinical evaluation was based on previous diagnosis, medical records, evaluation by a psychiatrist and/or Childhood Autism Rating Scale (CARS) and Raven's Progressive Matrices results. Whole exome sequencing was performed for two quartets and 10 trios (seven ASD/ID individuals and seven non ASD/ID DMDs) and rare exonic variants that reached minimum quality scores of coverage and allele balance were selected. Variant exclusion criteria included synonymous variants, nonsynonymous variants (CADD TRIO, *EBF3*, *CACNA1C* and *HEPACAM*. Two of them were classified as pathogenic variants based on ACMG guidelines and literature data (*HEPACAM* and *EBF3*), while only one variant was found in the Non-ASD DMDs (*CDK13*). Preliminary, these data suggest that ID and/or ASD-DMD individuals present a higher number of damaging variants in genes associated with neurodevelopment, compared to Non-ASD DMD individuals, suggesting the relevance of the genetic background for ASD/ID in DMD.
Financial support: CEPID/ FAPESP.

PrgmNr 2629 - Effects of somatic mutations on cellular differentiation outcomes and outlier behaviour in iPSC models

View session detail

Author Block: **P. PUIGDEVALL COSTA**[1], J. Jerber[2], P. Danecek[2], G. Sebastian[2], M. E. Hurles[2], S. Castellano[1], H. Kilpinen[1,2,3]; [1]Univ. Coll. London, London, United Kingdom, [2]Wellcome Sanger Inst., Cambridge, United Kingdom, [3]Helsinki Inst. of Life Sci., University of Helsinki, Finland
Disclosure Block: **P. Puigdevall costa:** None.
Induced pluripotent stem cells (iPSC) are an important tool for disease modelling. In particular, they offer unique opportunities to study the cellular basis of developmental disorders, as cell types and lineages relevant to human neurogenesis can be accessed through differentiation. Combining pooled iPSC-based experiments with single-cell sequencing has helped to tackle some of the heterogeneity of in vitro differentiation systems and is a powerful design to compare wild-type and mutant cells. However, differentiation outcomes of individual cell lines remain variable and not well understood. We explored whether somatic mutations acquired during reprogramming might affect the capacity of iPSC lines to differentiate. We leveraged exome sequencing data from the HipSci Project, available for iPSC lines as well as the fibroblasts from which they were derived, and analysed differentiation outcomes from four published experiments using HipSci iPSCs: dopaminergic neurons (DN; n=352 lines), macrophages (n=102), sensory neurons (n=85) and endoderm (n=86). While the total burden of acquired somatic mutations was not associated with the ability of iPSCs to differentiate, we observed significant differences in burden at key developmental gene-sets (pi1>0.8), such as known developmental disorder genes (DDD, www.ddduk.org) and cancer-associated genes (COSMIC-Tier1), between lines that differentiated successfully or failed. We then analysed pooled scRNA-seq data from the DN differentiations (Jerber et al. 2021 Nat Genet), which includes cell-lines profiled at three time points (days 11, 30, 52). We found that cell-type composition differences between failed and successful lines were already observed at the progenitor stage (d11), yet again likely driven by expression differences in developmental gene sets modulating commitment towards DN. We also found that expression of such genes correlates with early progenitor abundance. In order to study the effects of known disease mutations, their effects need to be distinguishable from cell line outlier behaviour caused by other factors. To this end, we identified outlier cell-lines with abnormal patterns of differentiation in the DN single-cell dataset that did not show any shift on mutational burden compared to non-outlier cell lines. We are currently characterising the differentiation behaviour of seven iPSC lines with individual CRISPR-engineered knock-outs of known DDD genes included in the DN experiment. Our study constitutes the first attempt to assess the feasibility and constraints to detecting effects of pathogenic genetic variants in iPSC-based, pooled single-cell experiments.

PrgmNr 2630 - Pathogenic *SPTBN1* variants cause an autosomal dominant neurodevelopmental syndrome

View session detail

Author Block: **M. A. Cousin**[1], B. A. Creighton[2], K. A. Breau[2], R. C. Spillmann[3], E. Torti[4], S. Dontu[1], S. Tripathi[5], D. Ajit[2], R. J. Edwards[2], S. Afriyie[2], J. C. Bay[2], K. M. Harper[2], A. A. Beltran[2], L. J. Munoz[2], L. Falcon Rodriguez[2], M. C. Stankewich[6], I. Helbig[7], S. E. McKeown[8], SPTBN1 Clinical Consortium, Undiagnosed Diseases Network, Genomics England Research Consortium, A. S. Beltran[2], M. T. Zimmermann[9], B. Temple[2], S. S. Moy[2], E. W. Klee[1], K. Tan[10], D. N. Lorenzo[2]; [1]Mayo Clinic, Rochester, MN, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Duke Univ., Durham, NC, [4]GeneDx, Gaithersburg, MD, [5]Med. Coll. of Wisconsin, Elm Grove, WI, [6]Yale Univ., New Haven, CT, [7]Children's Hosp. of Philadelphia, Philadelphia, PA, [8]Childrenâ□□s Hosp. of Philadelphia, Philadelphia, PA, [9]Med. Coll. of Wisconsin, Milwaukee, WI, [10]Duke Univ. Med. Ctr., Durham, NC

Disclosure Block: **M.A. Cousin:** None.

SPTBN1 encodes Î²II-spectrin, the ubiquitously expressed member of the Î²-spectrin family that forms micrometer-scale networks associated with plasma membranes. Î²II-spectrin is abundantly expressed in the brain, where it is essential for neuronal development and connectivity. Previous studies have shown that mice deficient in neuronal Î²II-spectrin have defects in cortical organization, global developmental delay, dysmorphisms, and behavioral deficiencies. These phenotypes, while less severe, are observed in haploinsufficient animals, suggesting that individuals carrying heterozygous variants in this gene may also present with measurable compromise of neural development and function. In this study, we report the identification of *de novo SPTBN1* variants in 29 individuals as a cause of a neurodevelopmental syndrome characterized by global developmental, language and motor delays, mild to severe intellectual disability, autistic features, seizures, behavioral and movement abnormalities, hypotonia, and variable dysmorphic facial features. Twenty-eight unique variants were identified (one individual has two *de novo* variants in cis) of which 22 are missense, three are nonsense, and three are canonical splice-site variants. Missense variants in codons Gly205, Thr268, Arg411 and Arg1003 were identified in more than one individual. Using cell-based systems including heterologous cell lines, primary mouse cortical neuron culture, and human induced pluripotent stem cell (iPSC) lines reprogrammed from peripheral blood mononuclear cells (PBMCs), we show that these *SPTBN1* variants lead to loss-of-function, gain-of-function, and dominant negative effects that affect protein stability, disrupt binding to key protein partners, and disturb cytoskeleton organization and dynamics. We further investigated the molecular rationale for these observations using in silico molecular modeling. Lastly, we assessed Î²II-SpHet mice and the results suggest that Î²II-spectrin LOF impairs global development and has a selective impact on social motivation and reward that may contribute to the autistic features and social behavior impairments manifested in some affected individuals. Our studies define *SPTBN1* variants as the genetic basis of a neurodevelopmental syndrome, expand the set of spectrinopathies affecting the brain and neural development, and underscore the critical role of Î²II-spectrin in the central nervous system.

PrgmNr 2631 - Schizophrenia risk mediated by microRNA target genes in 22q11.2 deletion syndrome

View session detail

Author Block: **S. Ying**[1], Z. Zhang[1], R. K. Yuen[2], A. S. Bassett[3]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]The Hosp. for Sick Children, Toronto, ON, Canada, [3]Univ of Toronto, Toronto, ON, Canada
Disclosure Block: **S. Ying:** None.
Background: 22q11.2 deletion syndrome (22q11.2DS) is associated with an over 20-fold increase in risk for schizophrenia. The 22q11.2 deletion region contains *DGCR8*, encoding a protein in the miRNA biogenesis pathway; global miRNA dysregulation imparted by *DGCR8* haploinsufficiency may play a role in mediating increased risk for schizophrenia. There is also evidence that rare genome-wide copy number variants (CNVs) increase schizophrenia risk in 22q11.2DS, and in the general population. We hypothesized that, within 22q11.2DS, in those with schizophrenia, there would be an enrichment of miRNA target genes that are overlapped by rare genome-wide CNVs.
Methods: We derived experimentally supported miRNA target genes (n=8464) from DIANA-TarBase, and obtained data available on rare genome-wide CNVs that overlapped genes in 22q11.2DS. We compared results for the miRNA target genes overlapped by these CNVs in 218 individuals with 22q11.2DS, with and without schizophrenia.
Results: Consistent with our hypothesis, among the 22q11.2DS cohort, we found a higher proportion of subjects in the schizophrenia group to have one or more miRNA target genes overlapped by a rare CNV (odds ratio (OR)=2.29, p=0.004). The miRNAs whose target genes contributed most to this differential target gene burden included miR-17-5p, miR-124-3p, and miR-342-3p, miRNAs with previous evidence for differential expression in 22q11.2DS experimental models and/or in schizophrenia and related neuropsychiatric disorders. Gene set enrichment results supported a role for miRNA target genes that belonged to several sets previously implicated for schizophrenia in the general population, including targets of FMRP and post-synaptic density (PSD)-related, when compared to all genes in the genome (FDR**Conclusion**: These data are the first to support a possible synergistic disruption of gene expression involving the target genes of dysregulated miRNAs and genome-wide rare CNVs as part of the complex mechanisms that increase risk for schizophrenia in 22q11.2DS. A comparable mechanism may be relevant to individuals with idiopathic schizophrenia and could therefore have implications for proposed novel miRNA-based therapeutics.

PrgmNr 2632 - Telomere length analysis in amyotrophic lateral sclerosis using large-scale whole genome sequence data

View session detail

Author Block: **A. Al Khleifat**, The Project MinE Consortium; Kingâ□□s Coll. London, London, United Kingdom
Disclosure Block: **A. Al Khleifat:** None.
Background Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease characterized by the loss of upper and lower motor neurons, leading to progressive weakness of voluntary muscles, with death following from neuromuscular respiratory failure, typically within 3 to 5 years. There is a strong genetic contribution to ALS risk. In 10% of cases or more, a family history of ALS or frontotemporal dementia is obtained, and the Mendelian genes responsible for ALS in such families have now been identified in about 80% of cases. Only about 14% of apparently sporadic ALS is explained by known genetic variation, suggesting that other forms of genetic variation are important. Telomeres maintain DNA integrity during cellular replication, differ between sexes, and shorten naturally with age. Gender and age are risk factors for ALS and we therefore investigated telomere length in ALS.
Methods: Samples were from Project MinE, an international ALS whole genome sequencing consortium that includes phenotype data. For validation we used donated brain samples from motor cortex from people with ALS and controls. Ancestry and relatedness were evaluated by principal components analysis and relationship matrices of DNA microarray data. Whole genome sequence data were from Illumina HiSeq platforms and aligned using the Isaac pipeline. We estimated telomere length by applying a bioinformatics analysis to the data. We tested the association of telomere length with ALS and ALS survival. **Findings**: There were 6,580 whole genome sequences, reducing to 6,195 samples (4,315 from people with ALS and 1,880 controls) after quality control, and 159 brain samples (106 ALS, 53 controls). Accounting for age and sex, there was a 20% (95% CI 14%, 25%) increase of telomere length in people with ALS compared to controls ($p = 1.1 \times 10^{-12}$), validated in the brain samples ($p = 0.03$). Those with shorter telomeres had a 10% increase in median survival ($p = 5.0 \times 10^{-7}$). Although there was no difference in telomere length between sporadic ALS and familial ALS ($p=0.64$), telomere length in 382 people with ALS due to expanded *C9orf72* repeats was less than in those without expanded *C9orf72* repeats ($p = 5.0 \times 10^{-4}$). **Interpretation:** Although telomeres shorten with age, longer telomeres are a risk factor for ALS and worsen prognosis.
Conclusions: It is likely that longer telomeres increase risk for ALS.

PrgmNr 2633 - Understanding the role of CLP1 in mammalian mRNA transcription and cleavage in neurodegeneration

View session detail

Author Block: **G. LaForce**[1], J. Farr[1], C. Akesson[1], O. Pinkard[1], T. Sweet[1], P. Ji[2], J. Coller[3], E. Wagner[2], A. E. Schaffer[1]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Univ. of Texas Med. Branch, Galveston, TX, [3]Johns Hopkins Univ., Baltimore, MD

Disclosure Block: **G. LaForce:** None.

Tight regulation of mRNA isoform expression is essential for neuronal development, maintenance, and function; however, the factors governing this process are largely unknown. We show the RNA kinase, CLP1, regulates mRNA isoform diversity through suppression of proximal transcription termination to promote expression of long mRNA isoforms. We generated human stem cell-derived motor neurons with CLP1 knockout, the motor neuron disease-associated CLP1 p.R140H variant, and the kinase-deficient CLP1 p.K127A variant, and performed transcriptome analyses. We identified distinct patterns of polyadenylation site usage resulting in imbalanced mRNA isoform expression of long genes important for neuronal development and function. We show similar mRNA misprocessing signatures in the CLP1 knockout and CLP1 p.K127A that differ from the CLP1 p.R140H, providing evidence that the disease-associated p.R140H allele exhibits toxic gain-of-function properties and kinase activity is required for CLP1 function in mRNA 3â□□-end processing. Together, these results identify a role for CLP1 in regulating mRNA isoform diversity and balance, of which perturbation causes broad transcriptional dysregulation and neurodegeneration.

PrgmNr 2634 - Wnt/beta-catenin pathway and cell adhesion deregulation in CSDE1-related intellectual disability and autism spectrum disorders

View session detail

Author Block: **E. El Khouri**[1], J. Ghoumid[2], D. Haye[3], F. Giuliano[4], L. DrÃ©villon[5], A. Briand-Suleau[6], P. De La Grange[7], V. Nau[8], T. Gaillon[9], T. Bienvenu[10], H. Jacquemin-Sablon[11], M. J. Goossens[9], S. Amselem[12], I. Giurgea[13]; [1]Hosp. Armand-Trousseau, Paris, France, Paris, France, [2]Lille Univ., Lille, France, [3]Ctr. Hosp.o-Univ.ire de Nice, Nice, France, [4]Ctr. Hosp.o-universitaire de Nice, Nice, France, [5]Hosp. Henri Mondor, CRETEIL, France, [6]Hosp. H. Mondor, AP-HP, Creteil, France, [7]GenoSplice, Paris, France, [8]HÃ´pital Trousseau, Paris, France, [9]Groupe Hosp.ier Henri Mondor, CrÃ©teil, France, [10]INSERM - Inst. Cochin, Paris, France, [11]INSERM, Paris, France, [12]Hosp. Armand-Trousseau, Paris, France, [13] Hosp. Armand-Trousseau, Paris, France, Paris, France

Disclosure Block: **E. El Khouri:** None.

Among the genetic factors playing a key role in the etiology of intellectual disabilities (IDs) and autism spectrum disorders (ASDs), several encode RNA-binding proteins (RBPs). In this study, we deciphered the molecular and cellular bases of ID-ASD in a patient followed from birth to the age of 21, in whom we identified a de novo *CSDE1* (Cold Shock Domain-containing E1) nonsense variation. CSDE1 encodes an RBP that regulates multiple cellular pathways by monitoring the translation and abundance of target transcripts. Analyses performed on the patient's primary fibroblasts showed that the identified *CSDE1* variation leads to haploinsufficiency. We identified through RNA-seq assays the Wnt/Î²-catenin signaling and cellular adhesion as two major deregulated pathways. These results were further confirmed by functional studies involving Wnt-specific luciferase and substrate adhesion assays. Additional data support a disease model involving APC Down-Regulated-1 (APCDD1) and cadherin-2 (CDH2), two components of the Wnt/Î²-catenin pathway, CDH2 being also pivotal for cellular adhesion. Our study, which relies on both the deep phenotyping and long-term follow-up of a patient with CSDE1 haploinsufficiency and on ex vivo studies, sheds new light on the CSDE1-dependent deregulated pathways in ID-ASD.

PrgmNr 2635 - An epigenome-wide view of osteoarthritis in primary tissues

View session detail

Author Block: **P. Kreitmaier**[1], M. J. Suderman[2], L. Southam[1], R. C. de Almeida[3], H. Konstantinos[1], I. Meulenbelt[3], J. Steinberg[4,1,5], C. Relton[2], J. M. Wilkinson[6], E. Zeggini[1,7]; [1]Inst. of Translational Genomics, Helmholtz Zentrum Muenchen, Neuherberg, Germany, [2]MRC Integrative Epidemiology Unit, Population Hlth.Sci., Univ. of Bristol, Bristol, United Kingdom, [3]Dept. of BioMed. Data Sci., Section Molecular Epidemiology, Leiden Univ. Med. Ctr., Leiden, Netherlands, [4]Cancer Council NSW, Woolloomooloo, Australia, [5]Sch. of Publ. Hlth., Faculty of Med. and Hlth., Univ. of Sydney, Sydney, Australia, [6]Dept. of Oncology & Metabolism, The Univ. of Sheffield, Sheffield, United Kingdom, [7]TUM Sch. of Med., Technical Univ. of Munich and Klinikum Rechts der Isar, Muenchen, Germany
Disclosure Block: **P. Kreitmaier:** None.
Osteoarthritis (OA) is a complex disease clinically characterized by chronic joint pain, stiffness and progressive deformity. Affecting more than 40% of people over the age of 70, OA contributes substantially to public health burden. Yet in spite of its high prevalence, there is no cure, and current treatment methods are limited to pain management and joint replacement. To promote the development of novel, efficient therapeutic approaches, it is first necessary to unravel the genomic architecture of osteoarthritis. Here, we investigate matched genotype and methylation profiles from macroscopically intact (low-grade) and degraded (high-grade) osteoarthritis cartilage, as well as synovium tissue from a total of 98 knee osteoarthritis patients undergoing joint replacement surgery.We conducted an epigenome-wide association study (EWAS) comparing matched low-grade and high-grade osteoarthritis cartilage within the same individual to identify methylation markers of cartilage degeneration. We further constructed random forest-based classifiers of cartilage degeneration. We generated genome-wide cis-methylation QTL (mQTL) maps in synovium, low-grade, and high-grade osteoarthritis cartilage. We used Mendelian randomization (MR) and colocalisation to identify epigenetic mechanisms mediating the effects of genotype on disease risk.The EWAS discovered 15,328 differentially methylated sites and 2,477 regions linked to cartilage degeneration. These regions were mapped to genes enriched in 76 Gene Ontology terms including a potential novel process linked to epithelium development. Using machine learning approaches, we derive methylation models of cartilage degeneration, which we validate with 82% accuracy in independent data. The mQTL analysis revealed widespread associations between genetic variants and methylation in all three examined joint tissues. Furthermore, mQTLs colocalised with 18 osteoarthritis-associated genetic loci, potentially elucidating their regulatory targets. Applying MR identified 19 methylation loci with a putative causal influence on osteoarthritis.In this study, we identify widespread epigenomic profile changes related to cartilage degeneration. We provide the first genome-wide map of mQTLs in knee OA affected tissues. This map enabled us to identify likely effector genes of GWAS signals, thus enhancing our understanding of the mechanisms underpinning disease development and progression in OA.

PrgmNr 2636 - Concurrence of Incontinentia pigmenti and thrombophilia increases the risk of Recurrent Early Pregnancy Loss

View session detail

Author Block: **M. V. Ursini**[1], E. Spinosa[1], M. B. Lioi[2], C. Casale[1], A. Pescatore[1], F. Fusco[1]; [1]IGB ABT CNR, Naples, Italy, [2]Univ. of Basilicata, Potenza 85100, Italy, Italy
Disclosure Block: **M.V. Ursini:** None.
Pregnancy loss, also referred to as miscarriage or spontaneous abortion, is generally defined as a nonviable intrauterine pregnancy up to 20 weeks of gestation. Here, we report the observation of recurrent miscarriages occurred in females with two diseases: Incontinentia Pigmenti (IP, MIM308300), a rare X-linked dominant disorder, and thrombophilia, an inherited disorder of blood clotting. The miscarriages are associated to both IP disease and thrombophilia. In IP the pregnancy loss in male fetus is due to the *NEMO/IKBKG* Loss of Function (LoF) mutations. In the thrombophilia it is due to hypercoaguable state which leads to arterial and/or venous thrombosis at the site of implantation or in the placental blood vessels.Three cases of IP females also suffered by thrombophilia. Two showed an history of Recurrent (>4) Early Pregnancy Loss (REPL) of both male and female fetus at less than 20 weeks of gestation. Each carried a different LoF mutation in exon 5 of *NEMO/IKBKG* gene (*c.523dup; c.646del; c.628_651insTG*). In addition, they carried mutations in genes of thrombophilia (Factor V Leiden_*FVL*, *MTHFR* mutation_*c.677C>T*, a prothrombin mutation_*G20210A* or *prothrombin II_PTII* mutation, or a protein S and/or C deficiency). Maternal comorbidities increase the risk of pregnancy loss. Thus we hypothesize that the co-occurrence of these two diseases in the same patient increases the risk of REPL. Here, we describe the IP and thrombophilic conditions and we perform genotype/phenotype correlation based on the genomic information of each patient to reveal shared genes and underlying functional systems. All patients belong to Incontinentia Pigmenti Genetic Biobank (IPGB; http://www.igb.cnr.it/ipgb/).

PrgmNr 2637 - Neonatal onset and severe complicated phenotype caused by heterozygous or homozygous new *ATL1* gene mutations

View session detail

Author Block: **A. Hamamie-Chaar**[1], M. Renaud[2], A-L. Bruel[3], C. Philippe[4], G. Banneau[5], V. Des Portes[6], N. Hadouiri[7], Q. THOMAS[8], L. Lambert[2], A. Hocquel[2], C. Thauvin-Robinet[9], L. Faivre[10]; [1]Pediatric PMR Dept., CHU Dijon, Dijon, France, [2]Service de GÃ©nÃ©tique Clinique HÃ´pitaux de Brabois - HÃ´pital d'enfants - CHRU de Nancy, Nancy, France, [3]Team EA4271 GAD, UniversitÃ© de Bourgogne, Dijon, Bourgogne, France, [4]Laboratoire de gÃ©nÃ©tique chromosomique et molÃ©culaire CHU de Dijon, Dijon, France, [5]UnitÃ© de NeurogÃ©nÃ©tique MolÃ©culaire et Cellulaire Ctr. de gÃ©nÃ©tique molÃ©culaire et chromosomique HÃ´pital de la PitiÃ©-SalpÃªtriÃ¨re APHP, Paris, France, [6]Service de neurologie pÃ©diatrique HÃ´pital Femme MÃ¨re Enfant CHU de Lyon HCL - GH Est, Lyon, France, [7]Service de MÃ©decine Physique et de RÃ©adaptation CHU Dijon, Dijon, France, [8]Dijon Burgundy Univ. Hosp., Dijon, France, [9]FHU-TRANSLAD, Dijon, France, [10]Hosp d'enfants CHU Dijon, Dijon, France
Disclosure Block: **A. Hamamie-Chaar:** None.

Hereditary spastic paraplegias (HSP) are clinically and genetically heterogeneous group of neurodegenerative diseases, characterized by a progressive spasticity of the lower limbs. Spastic paraplegia (SPG) A3 is the second most common form, mainly associated with an early age of onset, before the end of the first decade, and a pure phenotype. SPG3A is caused by heterozygous variants in *ATL1*, a gene which encodes atlastin GTPase 1.

One article has mentioned that autosomal recessive inheritance might exist.Indeed, Khan et al (2014) reported a large consangineous pakistani family with the segregation of a homozygous novel missense variant p.Arg118Gln in *ATL1* in six family members, affected by pure HSP, with no clinical signs in seven heterozygous carriers. We describe here two patients with an unusually neonatal onset and complex phenotype.

Patient 1 is a 6-year-old boy with severe axial hypotonia, dystonia and moderate spastic quadriplegia with a *de novo ATL1* mutation p.Lys406Glu detected by exome sequencing. He could not hold his head, he could not sit unassisted or walk, mobility required a wheelchair but he cannot use it himself. He also had swallowing disorders and was non-verbal. Intellectual abilities were difficult to assess was not affected. Patient 2 is a 7-year-old boy with spastic diplegia, cognitive impairment and epilepsy. He was not stable sit at 4yo, and he currently can only make few steps with human help. HSP panel gene revealed a homozygous novel mutation p.Arg403Glu. Probandâ□□s parents are consanguineous, they are asymptomatic and both carrier of the mutation at a heterozygous state.

In the literature, such severe clinical presentations are extremely rare. Regarding heterozygous patients, 3 mutations in *ATL1* (i.e., p.Pro344Ser, p.Met408Thr and p.Gly409Asp) resulted in very early-onset and severe complicated SPG3A. These codons were located in the linker or three-helix bundle region of atlastin 1. This part of the protein provides a structural basis for conformational changes and dimerization in homotypic ER membrane fusion. The patient 2 gives arguments in favor of a potential autosomal recessive inheritance, even if other patients should be described to confirm this hypothesis.

In conclusion, we report two further patients with neonatal onset HSP with complex phenotypes, both with heterozygous or homozygous *ATL1* variants. Further clinical descriptions will help defining the full clinical spectrum and mode of inheritance of *ATL1*-related phenotypes.

PrgmNr 2638 - Perturbation in hepatic lipid metabolism associated with altered TM6SF2 gene expression in HuH-7 cells

View session detail

Author Block: **A. Pant**[1], Y. Chen[1], A. Kuppa[1], X. Du[1], B. Halligan[1], E. Speliotes[1,2]; [1]Div. of Gastroenterology and Hepatology, Univ. of Michigan, Ann Arbor, MI, [2]Dept. of Computational Med. and Bioinformatics, Univ. of Michigan Med. Sch., Ann Arbor, MI

Disclosure Block: **A. Pant:** None.

Nonalcoholic fatty liver disease (NAFLD) is caused by accumulation of excess lipids in hepatocytes and can cause liver damage. Although NAFLD has become the leading cause of liver disease worldwide, few effective treatments exist for NAFLD making it a large unmet medical need. Genome wide association studies have identified strong association of NAFLD with non-synonymous E167K amino acid mutation in transmembrane 6 superfamily member 2 (TM6SF2) protein. The E167K mutation affects TM6SF2 stability and its carriers display increased hepatic lipids levels and lower serum triglycerides. While similar phenotype is evident in mice with TM6SF2 knockdown, effects of TM6SF2 on hepatic lipid metabolism is not completely understood. Here, we overexpressed wild-type or E167K variant of TM6SF2 or knocked down TM6SF2 expression in lipid-treated Huh-7 cells. We used biochemical assays, untargeted lipidomic analysis, RNAseq transcriptome analysis and high-throughput fluorescent imaging to determine changes in lipid metabolism and investigated molecular function of TM6SF2 in hepatocytes. Both knockdown and E167K overexpression increased acylglyceride levels which was decreased by wild-type TM6SF2 overexpression. Characterization of lipid droplets profile through high-content image analysis showed that overexpression of the E167K variant and TM6SF2 knockdown similarly affects lipid droplet phenotype to significantly increase the mean intensity while wild-type TM6SF2 had no effects. Although we did not see changes across lipid classes, we observed lipid chain remodeling for acylglycerides by TM6SF2 knockdown leading to a relative increase in species with shorter and more saturated side chains. Importantly, TM6SF2 knockdown and overexpression of wild-type TM6SF2 lead to significant changes in the abundance of several lipid species, including phosphoplipids, lysophospholipids, and acylglycerides. RNA sequencing revealed differential expression of several lipid metabolizing genes, including genes belonging to AKR1 family and lipases, primarily in cells with TM6SF2 knockdown. Taken together, our data shows that overexpression of TM6SF2 gene or its loss-of-function changes hepatic lipid species composition and expression of lipid metabolizing genes. Further, overexpression of E167K variant and TM6SF2 knockdown similarly increased hepatic lipid accumulation and lipid droplets profile further confirming a loss-of-function effect for variant. These results help us to understand how TM6SF2 and the E167K variant protect and promote NAFLD.

PrgmNr 2639 - Phenotypic Consequences of Variation in Elastin

View session detail

Author Block: **M. S. Williams**[1], A. Justice[1], M. A. Kelly[1], J. D. Green[1], R. Zaidi[1], B. A. Kozel[2]; [1]Geisinger, Danville, PA, [2]NIH, Bethesda, MD
Disclosure Block: **M.S. Williams:** Royalty(ies)/Honoraria; Prevention Genetics.
Introduction: Elastic fibers provide recoil to tissues that stretch, including the skin, blood vessels, and lungs. Haploinsufficiency for elastin (*ELN*) causes supravalvar aortic stenosis (SVAS OMIM#185500), characterized by arterial narrowing and stiffness, as well as hypertension, while missense changes in the C-terminal part of the gene lead to autosomal dominant cutis laxa (OMIM#123700). We hypothesize that variation in *ELN* may contribute more widely to connective tissue disease in the general population than is currently appreciated. Methods: Participants in the MyCode Community Health Initiative with exome sequence data were analyzed for common and rare sequence variation within *ELN*. ClinVar was interrogated for variants identified. In silico predictors were used to further identify variants of potential interest. Participants with variants of interest underwent dual chart review using a standardized abstraction tool that encompassed phenotypic findings in relevant organ systems (e.g. cardiovascular, pulmonary, renal). Reviewers were blinded to variant type. A Phenome-wide Association Study (PheWAS) was conducted with the participants carrying an *ELN* variant defined as cases and participants without an *ELN* variant defined as controls. Standardized and validated phenotypes (PheCodes) were developed and used for the analysis of electronic health records. Results: 115 living participants with a variant in *ELN* were identified from 145,000 MyCode participants. Roughly half of these had a variant (NM_000501.4:c.659C>T (p.Pro220Leu)) that has been annotated in ClinVar as likely pathogenic based on segregation of SVAS with the variant in several families. The other half had variants that were predicted to impact splicing. Chart review did not identify any patients with a diagnosis of SVAS or cutis laxa. 30 of the 115 participants (26%) had findings potentially consistent with elastinopathy, mostly related to vascular aneurysm. Conclusions: Preliminary analysis suggests that variation in *ELN* contributes to disease affecting elastic tissue function beyond classic phenotypes. Further studies will include expanding sequence analysis to 180,000 participants, PheWAS, enhancing variant annotation, functional variant analysis, and deep phenotyping.

PrgmNr 2640 - Saturation-scale functional evidence supports clinical variant interpretation in Lynch Syndrome

View session detail

Author Block: **A. Scott**[1], F. Hernandez[2], C. Horton[2], A. C. Chamberlin[2], X. Jia[1], S. Jayakody[1], B. Bharathi Burugula[1], T. Pesaran[2], E. C. Chao[2], R. Karam[2], J. Kitzman[1]; [1]Dept. of Human Genetics, Univ. of Michigan, Ann Arbor, MI, [2]Ambry Genetics, Aliso Viejo, CA
Disclosure Block: **A. Scott:** None.
Lynch Syndrome (LS) is a cancer predisposition syndrome affecting more than 1 in every 300 individuals worldwide. Tumorigenesis in LS is driven by germline loss-of-function variants in DNA mismatch repair (MMR) genes. Clinical genetic testing for LS can be life-saving but is complicated by a large burden of variants of uncertain significance (VUS), especially missense changes. We previously applied deep mutational scanning to measure functional effects for >94% of the 17,746 possible missense variants in the key LS gene *MSH2* (Jia et al, *AJHG*, 2021). Here, to establish the clinical validity of these functional data, and demonstrate their utility in large-scale variant reclassification, we overlaid them on clinical databases comprising >15,000 individuals with MMR gene variants detected by paired tumor and germline testing at a clinical genetic testing laboratory. To gauge the strength of evidence provided by our functional data, we curated a list of *MSH2* germline missense variants previously classified as Pathogenic (N=23) or Benign (N=26) as controls. To avoid redundant application of evidence, these validation variantsâ classifications were derived without use of any functional data. Our functional measurements agreed with the clinical interpretation for all 49 variants. Following recommendations for application of the functional evidence criterion using the ACMG/AMP variant interpretation framework (Brnich et al, *Genome Med*, 2019), our *MSH2* function map can be used with PS3/BS3 evidence codes given its observed strong concordance with these prior clinical interpretations. We next identified 720 standing VUS missense variants in *MSH2*. While most of these are predicted to have a neutral impact on gene function, 29 (4.0%) scored as deleterious in our function map, consistent with previously published rates among other cancer predisposition genes. We are pursuing reclassification for these variants, combining functional evidence with family history and tumor characteristics consistent with Lynch Syndrome. Additionally, in addition to retrospective VUS reclassification, to date these functional data have enabled resolution of five newly detected missense *MSH2* VUSs. High-throughput assays for mismatch repair loss of function provide a scalable method for VUS resolution, and serve as strong evidence criteria for variant classification.

PrgmNr 2642 - Spatial analyses of complement factor genes in AMD: Clarifying known genetic relationships and creating prediction models

View session detail

Author Block: **E. Palmer**[1], S. de Jong[2], M. Grunin[1], B. Jin[1], D. C. Rinker[3], C. Moth[3], J. A. Capra[4], W. S. Bush[5], J. L. Haines[5], A. den Hollander[6], IAMDGC; [1]Case Western Reserve Univ., Cleveland, OH, [2]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [3]Vanderbilt Univ., Nashville, TN, [4]Univ. of California San Francisco, San Francisco, CA, [5]Case Western Reserve Univ, Cleveland, OH, [6]Radboud Univ Nijmegen MC, Nijmegen, Netherlands

Disclosure Block: **E. Palmer:** Salary/Employment; CVS/Aetna.

Age-related macular degeneration (AMD) is the leading cause of blindness worldwide. Variants in complement factor genes are known risk factors for AMD, and potential therapeutic targets. Therefore, identifying if a given patient has a functional variant in one of these genes could aide in disease treatment stratification. Here, we incorporated protein 3D spatial context into interpretation of rare missense variants from the HRC-imputed International AMD Genomics Consortium (IAMDGC) dataset, containing 16,144 cases and 17,832 controls. We applied our protein spatial statistics algorithm (PathProx) to characterize variants that are more proximal to known AMD risk variants than observed presumed neutral variants from gnomAD. To evaluate the effect of rare missense variants on protein stability, we also calculated each variantâ□□s predicted free energy change (ddG). Variants proximal to known AMD variants or with a ddG>|2| were retained for gene-based testing using seqMeta. We also performed an orthogonal protein-based test that uses a structural kernel method (Pokemon) to identify spatial patterns of variants associated with case-status. Nominal p-values of 0.05 were considered significant in all tests. Finally, given the availability of protein expression data from cell *in vitro* experiments of *CFI* missense variants, we created a binary variable indicating if the variant caused a reduction in expression to C9, *CFH*, and *CFI*; ddG identified *CFI*, with *CFH* also suggestive of a possible association (p=0.051); Pokemon identified *C9*, *CFB*, *CFH*, and *CFI*. Finally, our model predicting a reduction in CFI protein had an area under the curve (AUC) of 0.76, thereby outperforming the model using CADD scores which had an AUC of 0.69. We have demonstrated that AMD risk variants in complement genes cluster within protein structures. This information may be useful in building predictive models that classify variants identified in AMD patients as benign or pathogenic. Further, we have shown that computationally predicted stability via ddG is a better indicator of a variantâ□□s potential to impact protein expression of CFI than CADD scores. Future work iterating bench validation and model training will allow us to identify which protein disruptions are most critical. This in turn will inform therapeutic development and enhance efforts to rapidly identify and validate the pathogenicity of novel variants identified in AMD patients.

PrgmNr 2643 - A network-based deep learning framework translates GWAS and multi-genomics findings to pathobiology and drug discovery for Alzheimerâ□□s disease

View session detail

Author Block: **J. Xu**[1], F. Cheng[1,2,3]; [1]Genomic Med. Inst., Lerner Res. Inst., Cleveland Clinic, Cleveland, OH, [2]Dept. of Molecular Med., Cleveland Clinic Lerner Coll. of Med., Case Western Reserve Univ., Cleveland, OH, [3]Case Comprehensive Cancer Ctr., Case Western Reserve Univ. Sch. of Med., Cleveland Clinic, OH

Disclosure Block: **J. Xu:** None.

Human genome sequencing studies have identified numerous loci associated with complex diseases, including Alzheimerâ□□s disease (AD). Translating human genetic findings (i.e., Genome-wide association studies [GWAS]) to biological mechanisms and therapeutic discovery, however, remains a major challenge. To address this critical problem, we present a **net**work **t**opology-based deep learning framework to identify disease-**a**ssociated **g**enes (NETTAG). NETTAG is capable of integrating multi-genomics data along with the protein-protein interactome to accurately infer likely risk genes and potential drug targets impacted by GWAS loci. Specifically, we leverage non-coding GWAS loci effects on expression quantitative trait loci (eQTLs), histone-QTLs, and transcription factor binding-QTLs, enhancer and promoter regions, open chromatin, and promoter flanking region from GTEx, NIH RoadMap and FANTOM5 databases.The key premises of NETTAG are that the disease risk genes: i) exhibit distinct functional characteristics compared to non-risk genes and therefore can be distinguished by their aggregated genomic features, and ii) converge to a limited number of pathobiological pathways captured by the human protein interactome. Via applying it to the latest AD GWAS data, we identified a set of likely risk genes (e.g., *APOE*, *BIN1*, *IL6* and *PICALM*) that are consistent with leading pathophysiological hypotheses of AD. NETTAG-predicted risk genes are: (1) significantly enriched in multiple AD-related pathobiological pathways; (2) more likely to be differentially expressed in single-nuclei RNA-sequencing data of AD patient brains in cell type-specific manners (including disease-associated microglia and astrocytes); and (3) are enriched in drug targets with available approved medications. In summary, our findings suggest that understanding of human pathobiology and therapeutic development could benefit from a network-based deep learning methodology that utilizes GWAS findings under the multimodal genomic analyses and the human protein interactome model. From a translational perspective, if broadly applied, the network-based deep learning tools could minimize the translational gap between genetic/genomic findings and precision medicine, is a significant challenge in genomic medicine.

PrgmNr 2644 - A pipeline for the identification of rare variants in multigenerational families with high prevalence of complex disorders

View session detail

Author Block: **E. RodrÃguez**, G. Chavarria-Soley, H. Raventos; Univ. de Costa Rica, San JosÃ©, Costa Rica
Disclosure Block: **E. RodrÃguez:** None.
Genome-wide association studies have been widely used for the identification of common variants associated with various diseases, however these variants only explain a small portion of the inheritance in complex diseases.
On the other hand, studies with extended pedigrees are valuable for the identification of novel rare variants that are transmitted to affected individuals. This rare variants are considered important factors for complex diseases.
However, many of the possible novel rare variants are usually filtered out by various tools whose filtering processes prioritize the permanence of known variants over unknown variants, minimizing the number of false positives but generating high rates of false negatives.
Here we aim to establish a practical pipeline that favors the identification of high-quality rare variants for germline variants calling in multigenerational families with high prevalence of complex disorders. The pipeline was tested using 114 WGS from members of a large multigenerational Costa Rican family with a high prevalence of complex psychiatric phenotypes. In addition, to reduce the needs of computational infrastructure, as well as execution times, a specific interval was applied in which the *Disrupted in schizophrenia 1* (*DISC1*) gene is located. This gene is found in a region in which a balanced chromosomal translocation t(1;11)(q42.1;q14.3) was previously reported linked to schizophrenia and psychosis in a Scottish family.
As a result, the pipeline allows to carry out a call for common and rare variants for genetic studies in multigenerational pedigrees with complex disorders, applying best practice procedures, genotypic refinement by pedigree and filtering of variants that prioritize the identification of high-quality rare variants.
Once the variants have been obtained and identified, they are available for later analysis in an association study.

PrgmNr 2645 - Aberrant splicing prediction across human tissues

View session detail

Author Block: **N. Wagner**[1], M. Ã□elik[2], F. HÃ¶lzlwimmer[1], V. YÃ©pez[1], C. Mertes[1], H. Prokisch[3], J. Gagneur[1]; [1]Technical Univ. of Munich, Garching, Germany, [2]Univ. of California Irvine, Irvine, CA, [3]Helmholtz Zentrum Muenchen, Neuherberg, Germany
Disclosure Block: **N. Wagner:** None.
Aberrant splicing is a major cause of genetic disorders but its direct detection in transcriptomes is limited to clinically accessible tissues such as skin or body fluids. While DNA-based machine learning models allow prioritizing rare variants, their performance on predicting tissue-specific aberrant splicing remains unknown. Here, we generated the first benchmark dataset for aberrant splicing prediction by applying the aberrant splicing caller FRASER [1] on 6,931 RNA-seq samples from 48 human tissues from the GTEx dataset, comprising 5 million rare variants in paired genotype data from 635 individuals. At 20% recall, state-of-the-art DNA-based models [2,3] cap at 2% precision. We constructed a tissue-specific splicing map (SpliceMap) by mapping and quantifying tissue-specific splice site usage genome-wide and modeling isoform competition. Using SpliceMap together with state-of-the-art DNA-based models increased precision by ten fold. Integrating RNA-sequencing data of clinically accessible tissues brought precision to 50%. Furthermore, we developed tissue-specific aberrant gene expression predictors with similar performance by leveraging that aberrant splicing often results in nonsense mediated decay. These findings, replicated in an independent cohort, substantially contribute to non-coding loss-of-function variant identification and have implications for non-invasive genetic diagnostics design and analytics.
[1] Mertes, Christian, et al. 'Detection of aberrant splicing events in RNA-seq data using FRASER.' Nature communications 12.1 (2021): 1-13.
[2] Cheng, Jun, et al. 'MMSplice: modular modeling improves the predictions of genetic variant effects on splicing.' Genome biology 20.1 (2019): 1-15.
[3] Jaganathan, Kishore, et al. 'Predicting splicing from primary sequence with deep learning.' Cell 176.3 (2019): 535-548.

PrgmNr 2646 - Analysis of 5â□□UTR Variation in Rare Disease Patients Reveals Variants of Potential Disease Relevance

View session detail

Author Block: **B. Bowles**[1], R. Olson[2], S. Senum[3], C. Schmitz[2], K. Clark[4], P. C. Harris[3], E. Klee[2]; [1]Mayo Clinic Dept. of Clinical and Translational Sci., Rochester, MN, [2]Mayo Clinic Ctr. for Individualized Med., Rochester, MN, [3]Mayo Clinic Polycystic Kidney Disease Discovery Lab., Dept. of Biochemistry and Molecular Biology, Rochester, MN, [4]Mayo Clinic Dept. of Biochemistry and Molecular Biology, Rochester, MN

Disclosure Block: **B. Bowles:** None.

Correct diagnosis is imperative to treating patients with idiopathic, suspected genetic conditions, yet current sequencing approaches leave up to 70% of these patients undiagnosed, significantly complicating both their medical care and genetic counseling. These remaining undiagnosed patients often suffer protracted diagnostic odysseys as they seek effective treatment. Variants within the 5â□□ untranslated region (UTR), particularly in upstream open reading frames (uORFs) that can act as regulatory elements, represent an especially under-examined source of disease. uORFs are short regions, typically 30bp - 600bp, that influence downstream gene translation through a variety of mechanisms, such as sequestering ribosomes or initiating nonsense-mediated decay of their associated mRNA. uORFs are present upstream of an estimated 54% of human transcripts. To better understand the range of human uORF variation and to evaluate the feasibility of uORF analysis as a part of clinical sequencing workflow, we retrospectively analyzed whole exome sequencing (WES) results from 469 idiopathic rare disease patients and 129 polycystic kidney disease (PKD) patients, as well as panel based sequencing results for an additional 979 PKD patients. We annotated sequencing files with variant interpretation information from a variety of sources, including population allele frequency, Combined Annotation Dependent Depletion (CADD) score, and evolutionary conservation score, and applied two deep learning tools to predict variant impacts on transcript ribosome load (TITER and FramePool). Our pipeline identified a median of 5 variants per patient that were predicted to have a deleterious impact on protein translational efficiency. These variants create new uORFs out of frame with the downstream coding sequence (CDS), are predicted to create N-terminal extensions in frame with the downstream CDS, or alter the function of existing uORFs. From these results, we have identified 5â□□UTR variants of uncertain significance upstream of *CTNND1*, *DPF2*, *HEPACAM*, *HSP90B1*, *NOTCH2*, *OTOF*, and *SMARCC2*, which have a potential role in initiating or modifying patient disease. This work demonstrates that analysis of 5â□□UTR variants can be incorporated into existing WES pipelines, and identifies a subset of variants with potential significance to patient disease. Further experimental evidence is necessary to ascertain the pathogenicity of these variants.

PrgmNr 2647 - Characterization of complex genomic rearrangements detected with long reads sequencing: two chromothripsis cases

View session detail

Author Block: **Y. Duffourd**[1], s. verdez[2], C. Martin[1], L. Faivre[3], C. Philippe[4], T. Christel[1], N. Chatron[5], S-B. Caroline[6], P. Callier[7], D. Sanlaville[8], A. Vitobello[9]; [1]CHU DIJON - INSERM U1231, DIJON, France, [2]CHU Dijon, Dijon, France, [3]Hosp d' Enfants, Dijon, France, [4]Ctr. Hosp.ier Univ.ire, Vandoeuvre les Nancy, France, [5]Service de GÃ©nÃ©tique, Hospices Civils de Lyon, Bron Cedex, France, [6]Service de GÃ©nÃ©tique, Hospices Civils de Lyon,, LYON, France, [7]CHU Le Bocage, Dijon, France, [8]HCL, CBPE, BRON Cedex, France, [9]CHU Dijon Bourgogne, Dijon, France
Disclosure Block: **Y. Duffourd:** None.
Background: Genome sequencing (GS) and especially long reads sequencing (LRS) become the efficient approach to identify complex rearrangements involving multiple breakpoints and low complexity regions. In order to test this method, we choose to study cases of chromothripsis, which is a brutal chromosome rearrangement involving multiple chromosomes. To characterize each complex events consecutive to chromothripsis, we choose to reconstruct the sequence of derivative chromosomes using a hybrid pipeline involving long reads and short reads sequencing.
Method: We deployed 30X short reads GS and 20X SMRT sequencing (Pacific Biosciences) to resolve complex SV in 2 patients with syndromic intellectual disability. Three others samples from same LRS batch were used for filtration. We established a new pipeline using information provided by frequency, depth and genomic phase to filter breakpoints identified by long-reads. We used Dnarrange[1] pipeline to analyze the position and sequences at breakpoints (BPs). Retained BPs were independently validated by PCR, Sanger sequencing, fluorescent in situ hybridization (FISH).
Results: We obtained 1794 and 1652 BPs for the first and the second patient respectively. After filtration we narrowed that list to 263 and 210 respectively. Dnarrange-link reconstructed 29 and 27 pieces of derivative chromosomes. We confirmed eight BPs by PCR and Sanger sequencing nine sequences involving 24 blocks with FISH.
Conclusions: This pipeline allowed us to filter, analyze and reconstruct the sequence of derivative chromosomes, highlighting 180 blocs involving chromosomes 3, 10, 12 for the first patient and 115 blocks involving chromosomes 4, 11, 13, 14, 15, 21 for the second patient. Overall the combination of short reads GS and LRS permitted to characterize structural variation at single nucleotide resolution with a small number of samples.
1 : Mitsuhashi, S., Ohori, S., Katoh, K. et al. A pipeline for complete characterization of complex germline rearrangements from long DNA reads. Genome Med 12, 67 (2020).

PrgmNr 2648 - Improved detection of allele-specific expression using Personalised ASE Caller (PAC)

View session detail

Author Block: **A. Saukkonen**[1,2], H. Kilpinen[3,2,4], A. Hodgkinson[1]; [1]Molecular and Med. Genetics, King's Coll. London, London, United Kingdom, [2]UCL Great Ormond Street Inst. of Child Hlth., Univ. Coll. London, London, United Kingdom, [3]Helsinki Inst. of Life Sci., Univ. of Helsinki, Helsinki, Finland, [4]Wellcome Sanger Inst., Cambridge, United Kingdom
Disclosure Block: **A. Saukkonen:** None.
A key finding in complex trait genetics is that most variants associated with disease are in non-coding regions of the genome and overlap significantly with variants modulating gene expression, thus making a strong case for the role of gene regulation in disease. On a population scale, variants influencing gene expression are identified through expression quantitative trait loci (eQTL) mapping, but this is not possible in cases where sample sizes are small, or variants are rare. For this reason, the detection of an imbalance in the expression of two alleles within a gene (allele specific expression, ASE) is potentially more informative about genes that are undergoing genetic regulation in *cis*, but detection of these events needs particular care, since technical features such as RNA-seq mapping biases can lead to false positives.
Here, we present a mapping and filtering pipeline, Personalised ASE Caller (PAC), that minimises errors in ASE detection through multiple steps incorporating better phasing of genetic data utilising RNA reads, parental genome mapping, reallocation of multi mapping reads, selecting optimal mapping parameters and incorporating haplotype level information. We apply PAC to simulated data and show that it assigns reads more accurately and in higher numbers than standard single genome mapping and other â□□gold standardâ□□ ASE methods, particularly in genomic regions close to indels and other genetic variants. We also use PAC to quantify ASE in population level data, using 670 whole blood samples from the GTEx project. Within these data, we find that ASE signals recapitulate eQTL effect sizes well, and the correlation between allelic fold change (aFC) measured from ASE and eQTL data is higher using PAC compared to standard mapping and other ASE methods. PAC is a standalone pipeline written in Nextflow and is available to use via https://github.com/anna-saukkonen/PAC

PrgmNr 2649 - MRSD: a novel quantitative approach to predict tissue-specific feasibility of RNA-seq in the investigation of splicing diversity

View session detail

Author Block: **C. Rowlands**[1,2], A. Taylor[2], G. I. Rice[1], N. Whiffin[3], H. Hall[4], G. C. M. Black[1,2], kConFab Investigators, R. T. O'Keefe[1], S. J. Hubbard[1], A. G. L. Douglas[5,6], D. Baralle[5,6], T. A. Briggs[1,2], J. M. Ellingford[1,2]; [1]Div. of Evolution & Genomic Sci., Sch. of Biological Sci., Faculty of Biology, Med. and Hlth., Univ. of Manchester, Manchester, United Kingdom, [2]Manchester Ctr. for Genomic Med., St Maryâ□□s Hosp., Manchester Univ. NHS Fndn. Trust, Hlth.Innovation Manchester, Manchester, United Kingdom, [3]Univ. of Oxford, Oxford, United Kingdom, [4]MRC Human Genetics Unit, Inst. of Genetics and Molecular Med., Univ. of Edinburgh, Edinburgh, United Kingdom, [5]Wessex Clinical Genetics Service, Princess Anne Hosp., Univ. Hosp. Southampton NHS Fndn. Trust, Coxford Rd, Southampton, United Kingdom, [6]Faculty of Med., Univ. of Southampton, Duthie Building, Southampton Gen. Hosp., Tremona Road, Southampton, United Kingdom

Disclosure Block: **C. Rowlands:** None.

Background: RNA-seq is a highly powerful technology allowing the examination of the transcriptomic landscape in individual tissues or cell types. It may, for example, allow discernment of key changes at the transcript level between cells in healthy and diseased states, or unstressed and stressed states. However, few resources exist to allow researchers to identify which genes and transcripts can be feasibly investigated in a given tissue.

Aims: To develop an approach that quantifies, for a given tissue, the experimental feasibility of using RNA-seq to investigate given genes or transcripts. Further, to develop an online resource through which these predictions can be accessed and visualized.

Methods: We leveraged publicly-available RNA-seq datasets generated through the Genotype-Tissue Expression (GTEx) project to derive a novel metric, the minimum required sequencing depth, or MRSD, for all human genes genome-wide. For a given tissue, MRSD represents the number of uniquely mapping RNA sequencing reads required to yield a user-specified level of coverage of a given transcript or splice junction.

Results: We evaluate that MRSD predicts transcript coverage from the RNA sequencing depth of a given sample with a high positive predictive value (90.1-98.2%, dependent on parameters). Through cross-referencing with human disease genes listed in the Genomics England PanelApp gene panel repository, we also demonstrate that, for 58.0% of Mendelian disease genes, a sequencing depth of 100 million RNA-seq reads or fewer is likely to be sufficient to observe the majority of pathogenic mis-splicing events in at least one of three tissues: whole blood, lymphoblastoid cell line (LCL) and skeletal muscle. Finally, we have created an online portal that allows users to rapidly query lists of genes or variants of interest and receive a prediction for the minimum RNA sequencing depth required to scrutinize their features of interest.

Conclusions: We have developed a robust predictive metric that will allow researchers to make informed decisions about the likely feasibility of RNA-seq-based analyses at the level of individual genes and transcripts, and across all manner of research questions. We anticipate that integration of MRSD scores into experimental planning will reduce redundant sequencing runs and improve the efficiency of transcriptomic investigations.

PrgmNr 2650 - Re-genotyping and tool combinations improve detection accuracy of Copy Number Variants in Whole Genome Sequencing data

View session detail

Author Block: **E. Iovino**, D. Paola, M. Seri, P. Tommaso; Univ. of Bologna, Bologna, Italy
Disclosure Block: **E. Iovino:** None.

Copy Number Variants (CNVs) are defined as the duplication (Dup) or deletion (Del) of DNA material larger than 50bp. Although CNVs are a clinically important class of genomic variation, their detection is known to be highly affected by different factors, such as the sequencing coverage, the size of the event and the underlying reference sequence structure (e.g. segmental duplications and/or DNA repeats). With the growing use of Whole Genome Sequencing (WGS) in the clinical setting, >70 tools for the detection of structural variants, including CNVs, have been developed; however, none of them achieves good call accuracy over the entire spectrum of CNV sizes and types (duplications/deletions), displaying false positive rates as high as 89%.The combination of different approaches is therefore indicated as a solution to increase accuracy, but there is currently no clarity on how several tools should be used together. Here, we evaluated the impact of combining different state-of-art tools and variant re-genotyping on achieving high call accuracy for high-coverage WGS data from 1KG NA12878 sample. To this end, we chose five tools (Delly, ERDS, Lumpy, Manta, Svaba) based on different detection methods (split-reads, read-depth, paired-reads, de-novo assembly) and previously assessed to be among the best performing approaches (Kosugi et al., 2019). We performed re-genotyping of CNVs with assigned PASS filter using SV2 and subsequently merged variant calls from each of the tools with SURVIVOR. We assessed False Positive (FP) and True Positive (TP) rates for re-genotyped as well as non re-genotyped variants called by each tool independently and for each of their combinations, using the call-set by Kosugi et al., 2019 as a gold-standard. We observed that successful re-genotyping (SRG;Dels=2998, Dups=487) improves power to discriminate between TP and FP calls both for deletions and duplications: 84% and 71 % SRG Dels (n=2534) and Dups (n=347), respectively were TP, against only 22 % Dels and 7% Dups with failed re-genotyping (FRG; Dels= 7169, Dups=9849). TP SRG rate increases to 90 % Dels and 80% Dups if large CNVs (> 50 kb) and each tool exclusive calls were excluded. As for FRG Del calls, but not Dups, TP rate rises to 83% by combining at least four tools. However, no difference was noticed between FRG CNVs larger and smaller than 50 kb. Our results suggest that SRG/FRG and tool combination may represent valuable parameters for increasing detection accuracy of small/medium (50 kb) CNVs identified in WGS data.

PrgmNr 2651 - A large-scale cohort metagenomic study of the Brazilian miscegenated population: building gut microbiome profiles of healthy and unhealthy individuals

View session detail

Author Block: **G. M. Novo-Filho**[1], G. F. S. Carvalho[1], E. A. Zanardo[1], G. S. Lemos[1], M. A. Zonta[1], G. A. Janaudis[1], L. D. Kulikowski[2,1]; [1]INSIDE RESEARCH AND DEVELOPMENT DIAGNOSTICS S.A., Sao Paulo, Brazil, [2]Univ. de Sao Paulo, Sao Paulo, Brazil
Disclosure Block: **G.M. Novo-Filho:** None.
The human organism hosts a wide variety of microorganism species that, through a mutualistic interaction, can provide biological advantages. The set of species of microorganisms that composes a given microbiota, what is called microbiome, is highly adaptable to nutrition and individual health status. Thus, microbiome profile studies can be used as a marker for quality of life and individual clinical condition. However, since the environment is one of the factors associated with the microbiome variability, population characteristics must be considered in individual and large-scale metagenomic studies. Thus, we sequenced the 16S gene (V3/V4 region) of fresh stools samples form 449 patients in a large-scale cohort metagenomic study, aiming to build a gut microbiome profile from healthy and unhealthy Brazilian population. Individuals with any clinical complaint or pre-diagnosed clinical condition, including diarrhea, dyspepsia, celiac disease, ulcerative colitis, Crohn's disease and others, were included in the non-healthy group (NH), while individuals with no complaints nor pre-diagnosed diseases were included in the healthy group (H). Thereafter, we compared the average representation of the great phylum in both groups. We identified a significant variability in the average proportion of 4 large phylum between the H and NH groups. The greatest variability occurred between the Firmicutes (F) and Bacteroidetes (B) phylum, with 2.10 of F/B ratio in the H group and 1.56 in the NH group. Furthermore, in the healthy group we identified a greater abundance of the Actinobacteria phylum in relation to the Verrucomicrobia phylum, whereas in the non-healthy group there is a greater abundance of Verrucomicrobia. Increased abundance of Proteobacteria phylum, which includes several inflammatory species, is broadly associated with an unhealthy microbiota, however, in this study, the abundance of this phylum between the H and NH groups was similar, suggesting that Brazilian population characteristics may influence increased abundances of bacteria in the Proteobacteria phylum even in healthy individuals. The average phylum representation in a gut microbiome study can be used as a marker for individual health status. In addition, metagenomic tools may allow a better follow-up of the clinical course and effectiveness in the treatment of patients with different clinical conditions.

PrgmNr 2653 - High yield depletion for an RNA-seq application using an innovative benchtop platform for workflow automation and miniaturization

View session detail

Author Block: **S. Aguilar**; INOREVIA, Paris, France
Disclosure Block: **S. Aguilar:** None.

Massively parallel sequencing of RNAs has unlocked the study of the transcriptome, enabling genome wide transcriptional profiling. While mRNAs are of great interest as they encode for proteins, they make up only 3-5% of total RNAs, with the majority (80%) comprising ribosomal RNAs (rRNAs). Therefore, RNA-seq applications face the challenge of efficiently enriching mRNA species prior to library preparation. To improve transcriptomic applications, Magelia, a disruptive automated multi-OMICs platform combines patented technologies for sample and reagent volume reduction, and efficient bead manipulation.Here, we have evaluated the advantages of using the platformâ□□s rRNA depletion application for RNA-seq. Equal amounts of human control total RNA were treated in parallel manually and in the Magelia for rRNA depletion using Illuminaâ□□s Ribo-Zero Plus rRNA Depletion Kit. After depletion, treatment in the platform allowed us to recover 3x more mRNAs. This impacted library preparation, where 10x more sequencing ready material was recovered for samples treated in the instrument. Importantly, sequencing revealed superior depletion for the instrument treated samples as 27,5x less rRNA reads were counted. As the control RNA used for the study was extracted from a well characterized cell line, the Harmonizome database was used to identify highly transcribed genes across studies. All top 20 expressed genes were identified in the top 15% of transcription levels, corroborating the biological validity of these data. Finally, more gene fusion events were detected in samples treated in the platform, emphasizing the improved level of enrichment achieved with this benchtop solution. This multi-OMICs platform provides unprecedented transcriptome resolution by improving rRNA depletion levels, this, with a 5x reagent volume reduction. Harnessing the strengths of its patented technologies, this depletion solution has also been recently applied to provide unprecedented resolution into a bat virome study in collaboration with researchers of the University of Chile. Fully automated RNA-seq library preparation is currently being validated downstream of this depletion application. This adds to the platformâ□□s ever-growing portfolio of NGS applications, including WGS library preparation with up to 50x DNA input reduction when compared to kitâ□□s limits. In addition to significant hands-on time reduction, the platform provides a smart combination of miniaturization and automation. This unique combination of benefits puts the platform in a unique position to become the leader in miniaturized sample preparation for modern biology and medicine.

PrgmNr 2654 - Integrating expression and methylation data into GWAS for the discovery of novel colorectal cancer risk loci

View session detail

Author Block: **C. Fernandez-Rozadilla**[1], M. Timofeeva[2], Z. Chen[3], P. J. Law[4], M. Thomas[5], S. Schmit[6], V. DÃez-Obrero[7], L. Hsu[8], Asia Colorectal Cancer Consortium (ACCC), Colon Cancer Family Registry (CCFR), Genetics and Epidemiology of Colorectal Cancer (GECCO), Study of Colorectal Cancer in Scotland (SOCCS), Transdisciplinary Studies of Genetic Variation ofColorectal Cancer (CORECT), Colorectal Cancer Genetics (COGENT), V. Moreno[7], G. Casey[9], S. B. Gruber[10], I. Tomlinson[11], W. Zheng[12], M. Dunlop[11], R. Houlston[13], U. Peters[14]; [1]Inst. de Investigacion Sanitaria de Santiago - IDIS, Santiago de Compostela, Spain, [2]Univ. of Southern Denmark, Odense, Denmark, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]The Inst. for Cancer Res., London, United Kingdom, [5]Fred Hutch, Redmond, WA, [6]Cleveland Clinic, Cleveland, OH, [7]Bellvitge BioMed. Res. Inst. (IDIBELL), Barcelona, Spain, [8]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [9]Univ. of Virginia, Charlottesville, VA, [10]USC, Los Angeles, CA, [11]The Univ. of Edinburgh, Edinburgh, United Kingdom, [12]Vanderbilt Univ Med Ctr., Nashville, TN, [13]The Inst. of Cancer Res., Sutton, United Kingdom, [14]Fred Hutchinson CA Res Ctr, Seattle, WA
Disclosure Block: **C. Fernandez-Rozadilla:** None.

Colorectal cancer (CRC) is the third most commonly diagnosed cancer. Despite recent success of genome-wide association studies (GWAS) and discovery of common genetic variants in over 150 loci associated with CRC risk, much of the heritable risk and the biology underpinning CRC susceptibility remain unexplained.

To gain additional insights into CRC genetics, we conducted a GWAS meta-analysis totalling 100,204 CRC cases and 154,586 controls of European (73% of total) and East Asian (27%) ancestry. We identified 46 new loci associated with CRC risk at PWe therefore integrated the gene expression and methylation data from 1,077 and 488 samples of normal colorectal mucosa from 6 in-house datasets plus the transverse colon data from the Genotype-Tissue Expression (GTEx) project v8 into GWAS summary statistics to perform transcriptome- (TWAS) and methylome-wide association studies (MWAS). TWAS analyses identified 136 candidate risk genes, of which 18 at 11 loci lay over 1Mbp away from a GWAS-identified risk SNP, and were thus considered novel. Conditional TWAS analyses showed these 11 loci to be independent of GWAS SNPs and identified a further 2 genes independent of a GWAS locus within 1Mb. Likewise, MWAS analyses identified 15 CRC-associated CpGs annotated to 11 new risk loci previously undiscovered by GWAS or TWAS.

TWAS and MWAS have successfully identified 24 novel CRC susceptibility regions, highlighting the potential of multiomic data for discovering missing heritability. Moreover, TWAS and MWAS also provide a list of candidate genes to test for causal effects on disease risk, and this may be helpful for identifying new relevant CRC molecular pathways. We identified novel candidate target genes linked to microbiome, insulin pathway, hyperlipidaemia, and prostaglandin metabolism, which could lead to new drug targets for chemoprevention.

PrgmNr 2655 - Lipid pathway signatures discriminate subgroups linked to race/ethnicity in metabolomic and proteomic analysis from the TOPMed Study

View session detail

Author Block: **M. Sevilla**[1], A. K. Manning[1], K. D. Taylor[2], TOPMed Diabetes Working Group, TOPMed Multi-Omics Working Group; [1]Massachusetts Gen. Hosp., Boston, MA, [2]The Lundquist Inst. for BioMed. Innovation, 1124 W. Carson Street, Torrance, CA 90502, Torrance, CA
Disclosure Block: **M. Sevilla:** None.
The advent of omics studies offers an unparalleled opportunity to understand the processes involved in the development of complex diseases. Insights from diverse study samples are necessary to understand health disparities but omics profiles correlated with race/ethnicity or ancestry could be caused by environmental exposures or differences in allele frequencies. The objectives of this study were to identify and describe metabolomic and proteomic profiles that correlate to race/ethnicity or ancestry in two multi-ethnic cohorts from the NHLBIâ□□s TOPMed Study. Studying 2,398 individuals with proteomic and metabolomic data from the Multi-Ethnic Study of Atherosclerosis (MESA) and Womenâ□□s Health Initiative (WHI) cohorts, we defined both cohort-reported race/ethnicity subgroups and ancestry subgroups derived from allele differences linked to continent-specific haplotypes. Our labels were African-American (AA), East Asian (EA), European (EU), and Hispanic (HS) for the cohort-reported race/ethnicity subgroups. We used Orthogonal Projection to Latent Structures of Discriminant Analysis (OPLS-DA) to identify the proteomic or metabolomic profiles that distinguished each subgroup from the rest of the sample. We then used pathway analysis for the biological classification of these signatures. The metabolomic profile of the AA subgroup was characterized by both high carbon-number and low double-bond Diacylglycerols (DAGs) and Triacylglycerolâ□□s (TAGs) without any significant pathways, whereas the EA subgroup was characterized by high carbon-number and high double-bond DAGs and TAGs implicating the glycerophospholipid pathway (q=0.03). The same pathway was highlighted in the EU subgroup (q=0.03). The HS subgroup profile was also characterized by Phosphatidylethanolamines (PEs). We found more significant metabolites in the MESA cohort compared to WHI, but remarkably almost all of the metabolites identified in the WHI cohort were identified in the MESA cohort. Proteomic signatures that showed good discriminatory capability and were unique to a subgroup were Ephrin-A4 in the AA subgroup, and Lymphotoxin b R and Ephrin-A5 in the EA subgroup, and Transgelin-2 in the EU subgroup. Our results suggest that lipid profiles comprised DAGs, TAGs, and PEs are signatures of our subgroups. Such signatures could be important to understand health disparities but also could be included in models linking genetic variants and omics values to complex diseases to leverage larger sample sizes and correct for potential confounding due to these subgroup signatures.

PrgmNr 2656 - MED13L knockout in cerebral organoids disrupts retinal gene expression regulation

View session detail

Author Block: **J. Ghoumid**[1,2], R. Ziffra[2], D. Laboy Cintron[2], J. Sige[1], N. Ahituv[3]; [1]ULR7364 RADEME, Univ. of Lille, Lille, France, [2]Dept. of Bioengineering and Therapeutic Sci., Univ. of California San Francisco,, San Francisco, CA, [3]Dept. of Bioengineering and Therapeutic Sci., Univ. of California San Francisco, San Francisco, CA
Disclosure Block: **J. Ghoumid:** None.
The Mediator is a large coregulator complex conserved from yeast to human, taking part in the recruitment of RNA Polymerase II. It has emerged as a master coordinator of development and cell lineage determination through interactions with various transcription factors and epigenetic regulators. Mediator Complex is organized into four modules, i.e. Tail, Middle, Head, and the kinase-module. In vertebrates, the kinase-module comprises four proteins: CDK8, CCNC, MED12, and MED13, or their respective paralogs: CDK19, MED12L, and MED13L. Rare variants of the genes encoding this module have been linked to human neurodevelopmental disorders.
We developed a *MED13L*-knockout cerebral organoids from hIPSc (human Induced Pluripotent cells). We characterized *MED13L* KO and wild-type cerebral organoids through gene expression and chromatin accessibility analysis at the single-cell level.
We found in wt cerebral organoids, development of mature cortical neurons of both upper and deep layers (*BCL11B*, *SATB2*), glutamatergic and GABAergic neurons with apparently functional synapses (*GRIA1*, *GRIA2*, *GABRB3*). In *MED13L* KO cerebral organoids, gene expression analysis revealed that *MED13L* knockout led to a retinal commitment, with robust expression of retinal markers (*RAX*, *VSX2*) and photoreceptors (*USH2A*). Combining single-cell chromatin accessibility data allowed identification of co-accessible pairs of DNA elements, connecting regulatory elements to their putative target genes. Comparatively to the wt cerebral organoids, we found in the *MED13L* KO cerebral organoids a larger number of accessible regulatory elements driving high expression of genes critical for retinal development, including *PAX6*, *NEUROD1*, and *VSX2*. Basing on these data, *MED13L* is probably critical in the negative control of early genes inducing a retinal fate. This mechanism is likely to be critical to allow a proper cortical commitment to the developing neurons.

PrgmNr 2657 - Same role but different actors in the genetic regulation of post-translational modification of two distinct proteins

View session detail

Author Block: **A. Landini**[1], I. TrbojeviÄ□-AkmaÄ□iÄ□[2], P. Navarro[3], Y. A. Tsepilov[4,5], S. Sharapov[4], F. Vuckovic[2], O. PolaÅiek[6,7], C. Hayward[3], T. Petrovic[2], M. Vilaj[2], Y. S. Aulchenko[4], G. Lauc[2,8], J. F. Wilson[1], L. Klaric[3]; [1]Ctr. for Global Hlth.Res., Usher Inst., Univ. of Edinburgh, Edinburgh, United Kingdom, [2]Genos GlycoSci. Res. Lab., Zagreb, Croatia, [3]MRC Human Genetics Unit, Inst. for Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom, [4]Lab. of Glycogenomics, Inst. of Cytology and Genetics, Novosibirsk, Russian Federation, [5]Lab. of Theoretical and Applied Functional Genomics, Novosibirsk State Univ., Novosibirsk, Russian Federation, [6]Dept. of Publ. Hlth., Sch. of Med., Univ. of Split, Split, Croatia, [7]Gen-info Ltd., Split, Croatia, [8]Faculty of Pharmacy and Biochemistry, Univ. of Zagreb, Zagreb, Croatia

Disclosure Block: **A. Landini:** None.

Post-translational modifications (PTMs) are ubiquitous mechanisms used by cells to diversify and extend their protein functions beyond what is dictated by the genome. While PTMs are known to be involved in regulating almost all cellular events, genetic regulation of PTMs themselves has not been extensively investigated. Protein glycosylation, one of the major PTMs, has been linked to the ageing process and a wide variety of diseases, ranging from rheumatoid arthritis, type 2 diabetes, Crohnâ□□s and Parkinsonâ□□s disease to cancer. Nevertheless, genetic regulation of glycosylation is yet not fully understood. In this study, we compared for the first time the genetic regulation of the same post-translational modification of two distinct proteins -glycosylation of transferrin and immunoglobulin G (IgG). By performing genome-wide association analysis of transferrin glycome, we identified 10 significantly associated loci (P MGAT5, ST3GAL4, B3GAT1; IgG - *MGAT3, ST6GAL1*) as well as shared associations (*FUT6, FUT8*). Colocalization analyses of the latter suggest that distinct causal variants in FUT genes regulate fucosylation of the two proteins. We propose that these variants affect the binding of different transcription factors in different tissues, with fucosylation of IgG being regulated by IKZF1 in B cells and fucosylation of transferrin by HNF1A in liver. Uncovering the genes responsible for transferrin glycosylation and comparing them with those underlying IgG glycosylation, our results are beginning to unravel the genetic regulation of glycosylation, one of the most common PTMs. We show that both unique and shared genes are involved in glycosylation of two proteins, and suggest that different underlying causal variants in the same gene are regulating glycosylation of distinct proteins in a tissue-specific manner.

PrgmNr 2658 - Understanding the genetic contribution to neuroinflammatory cell traits using iPSC-derived microglia

View session detail

Author Block: **M. Perez-Alcantara**[1], Y. Chen[1], S. Washer[2], J. Steer[1], D. Ebner[3], S. Cowley[2], G. Trynka[1], A. Bassett[1]; [1]Wellcome Sanger Inst., Hinxton, United Kingdom, [2]William Dunn Sch. of Pathology, Oxford, United Kingdom, [3]Target Discovery Inst., Nuffield Dept. of Med., Oxford, United Kingdom

Disclosure Block: **M. Perez-Alcantara:** None.

There is mounting genetic evidence, both from common and rare variants, that implicates microglia in neurodegenerative disease. Microglia perform various immunological functions in the brain as sentinels, housekeepers and mediators of neuroinflammatory response: these include cellular signalling, migration and phagocytosis. Our aim is to assess how these cellular functions might be perturbed in neurodegenerative disease (particularly Alzheimerâ□□s but also Parkinsonâ□□s, amyotrophic lateral sclerosis and frontotemporal dementia), and the genetic contribution to these phenotypes. To achieve this, we will examine the role of neurodegeneration-associated variants and loci in the gene expression and cellular phenotypes of iPSC-derived microglia differentiated *in vitro*. We will perform pooled differentiations towards microglia of iPSCs from ~250 donors from the human pluripotent stem cell initiative (hipSCi), as well as candidate gene KOs and disease relevant mutant lines on isogenic background. This will be followed by phenotyping (for functions such as phagocytosis and chemotaxis) and single cell RNA-seq (scRNA-seq) to assess gene expression. Finally, we will perform expression and phenotypic Quantitative Trait Locus (QTL) mapping and colocalization with common variants identified by genome-wide association studies of neurodegenerative traits, with the aim of linking these variants to their target genes.We are improving the *in vitro* microglia maturation protocols to achieve cells more phenotypically and transcriptomically similar to primary microglia. Phenotyping protocols are adapted for pooled designs (10-20 donors/pool) e.g. with FACS-based outcomes and whole genome sequencing followed by deconvolution of donor identity for phenotypic QTL analysis. Expression QTL analysis of pooled microglia (using established methods such as tensorQTL) will be performed on scRNA-seq from microglial subpopulations under resting and stimulated (e.g. with LPS) conditions, to map the transcriptomic changes that occur during immune response. We will also perform a targeted single cell CRISPR KO screen and phenotyping of known and candidate disease-relevant mutants highlighted by previous prioritization studies. Finally, a subset of phenotypic assays will be applied to a genome-wide CRISPR screening to identify more genes involved in microglial processes.

PrgmNr 2659 - Whole Blood RNA Sequencing in a Cohort of Undiagnosed Pediatric Patients

View session detail

Author Block: **H. Hou**[1], L. Kyriakopoulou[2,1], K. E. Yuki[1], S. Barnes[1,3], A. Ramani[4], A. Celik[4], M. Braga[2], M. Gloven-Brown[2], M. Brudno[5], S. S. Meyn[6], G. Costain[7], C. R. Marshall[2,8,9], A. Shlien[1,10], J. Dowling[1,3,11], M. D. Wilson[1,3]; [1]Genetics and Genome Biology, SickKids Res. Inst., Toronto, ON, Canada, [2]Genome Diagnostics, Dept. of Paediatric Lab. Med., The Hosp. for Sick Children, Toronto, ON, Canada, [3]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [4]The Ctr. for Computational Med., SickKids Res. Inst., Toronto, ON, Canada, [5]The Ctr. for Computational Med., The Hosp. for Sick Children, Toronto, ON, Canada, [6]Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, [7]The Hosp. for Sick Children, Toronto, ON, Canada, [8]The Ctr. for Applied Genomics, The Hosp. for Sick Children, Toronto, ON, Canada, [9]Ctr. for Genetic Med., The Hosp. for Sick Children, Toronto, ON, Canada, [10]Dept. of Lab. Med. and Pathobiology, Univ. of Toronto, Toronto, ON, Canada, [11]Dept. of Paediatrics, Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **H. Hou:** None.

Background: RNA sequencing (RNA-seq) has the potential to improve our ability to interpret the functional and clinical significance of the genetic variants identified by Whole Exome Sequencing (WES) and Whole Genome Sequencing (WGS). A cohort of 134 pediatric patients with heterogeneous phenotypes collected through the SickKids Genome Clinic (Centre for Genomic Medicine) underwent testing by gene panels, microarray, and subsequently WGS. This resulted in the identification of pathogenic or plausibly causative variants in 44% of the samples. In this study, we set out to assess the utility of blood RNA-seq for the diagnosis of patients with diverse clinical indications and develop an interpretation schema to identify candidate genes.
Methods: We developed a clinical-grade, automated, scalable, and robust end-to-end RNA-seq library preparation platform capable of processing up to 96 samples at one time. We generated whole-blood RNA-seq from 134 individuals using this automated RNA-seq platform. Using a bioinformatics pipeline that incorporates various published tools and custom scripts, we identified expression outliers, aberrant splicing events, and allele-specific expression. We then took a disease gene-centric interpretation approach to identify clinically relevant transcriptional aberrations.
Results: Despite samples being archived for >3 years, we were able to generate good-quality RNA-seq data from all samples. With a median sequencing depth of 115M, we detected >70% of genes in curated genes panels related to phenotypes observed in our cohort. We identified expression outliers and/or aberrant splicing events in clinically relevant genes in 23.1% (31/134) blood RNA-seq data from our patient cohort. Specifically, we found RNA evidence supporting previously identified genetic variants in 30.5% (18/56) patients and proposed candidate genes in 16.6% (13/78) cases with negative WES/WGS findings. For example, we identified a known Joubert syndrome gene CEP120 as a decreasing expression outlier that contained a deep intronic splicing alteration in a patient with global developmental delay, abnormality of brain morphology, and dysmorphic facial features.
Conclusion: Our results support the use of clinical blood RNA-seq to facilitate genome diagnostics in pediatric patients with diverse phenotypes.

PrgmNr 2660 - *de novo* variant calling identifies cancer mutation profiles in the 1000 Genomes Project

View session detail

Author Block: **J. Ng**[1], P. Vats[2], E. Fritz-Waters[3], E. Padhi[4], Z. L. Payne[1], S. Leonard[3], S. Sarkar[1], E. Sams[5], M. West[2], C. Prince[3], L. Trani[6], M. Jansen[3], G. Vacek[2], M. Samadi[7], T. T. Harkins[2], C. Pohl[3], T. N. Turner[1]; [1]Washington Univ. Sch. of Med., St. Louis, MO, [2]NVIDIA, Ann Arbor, MI, [3]Res. Infrastructure Services, Washington Univ. Sch. of Med., St. Louis, MO, [4]Washington Univ. in St Louis, St Louis, MO, [5]Washington Univ. in St. Louis, St. Louis, MO, [6]The Elizabeth H. and James S. McDonnell Genome Inst., St. Louis, MO, [7]NVIDIA, San Jose, CA

Disclosure Block: **J. Ng:** None.

Detection of *de novo*â¯variants (DNVs) is critical for studies of disease-related variation and mutation rates and have been shown to occur at 50 to 100 DNVs per individual per generation. In developing a GPU-based workflow to expedite DNV calling, we applied the 602 parent-child sequenced trios from the publicly available 1000 Genomes Project as a set of controls. The DNA that was used for whole-genome sequencing from these individuals was derived from lymphoblastoid cell lines (LCLs). We detected 445,711 total DNVs with a bimodal distribution of DNVs, having a first peak at approximately 200 DNVs per child and a second peak at 2,000 DNVs. While the DNA from these individuals may contain cell line artifacts, these artifacts are propagating additional DNVs every cell passage, implying that the controls are not static. Evidence to support this finding include reduction in DNVs at CpG sites and in percent of DNVs with a paternal parent-of-origin with increasing number of DNVs. Detailed assessment of individual NA12878 across multiple genome datasets from 2012 to 2020 reveals increasing number of DNVs over time. Mutation signature analysis across this control set revealed individuals had either 1) age-related, 2) B-cell lymphoma, or 3) no prominent signatures.â¯ Additionally, clinically-relevant DNVs associated with twelve diseases were identified in individuals in this cohort. We also applied the DNV GPU workflow to 4,216 parent-child trios from the Simons Simplex Collection (SSC) with DNA derived from blood. This was in contrast to the 1000 Genomes dataset with DNA from LCLs. When testing our approach on the SSC data, we saw the expected results confirming the quality of our DNV workflow. The average DNV count was 78 per individual in the 4,216 trios. We also saw that an average of 18.3% of the DNV sites were found at CpG sites across the trios, which is close to the expected 20%. Our approach allows for fast and accurate detection of DNVs, which is especially important when analyzing large disease datasets, such as when studying autism. Our approach also shows that cell line artifacts present in lymphoblastoid cell lines are not always random but rather are associated with cancer mutation profiles with important implications for use of the variant data as a control.

PrgmNr 2661 - Building a predictive matrix model for autism genotype-phenotype associations

View session detail

Author Block: **E. Sams**, T. N. Turner; Dept. of Genetics, Washington Univ. Sch. of Med., St. Louis, MO
Disclosure Block: **E. Sams:** None.
Characterizing the structure and function of the human genome in association with phenotypic presentation at the molecular and organismal level is critical for advancing our understanding of human health and disease. In this study, we constructed a predictive matrix model for genotype-phenotype associations using previously collected genotypic and phenotypic data. The genetic event used to build this model is the presence of a copy number variant (CNV) in the 16p11.2 region. Recurrent CNVs in this region are strongly associated with autism and other developmental disorder phenotypes. The relevant data was obtained from the Simons Foundation Autism Research Initiativeâ□□s (SFARI) Simons Searchlight Project cohort of families affected by a 16p11.2 CNV. This cohort contains probands as well as affected and/or unaffected family members and thus includes both *de novo* and inherited 16p11.2 CNVs. The genotype section of the matrix for each sample was constructed by listing the copy number for each of the 52 genes in the 16p11.2 critical region using coordinates for CNVs confirmed by two technologies (microarray and whole exome sequencing). Every sample in the matrix also received a score indicating if they are diagnosed with autism or other related developmental disorder phenotypes, including coordination disorder, intellectual disability, and behavioral disorders. In addition to this 16p11.2 model, a genome-wide matrix was built in which the copy number for every protein coding gene in the genome was recorded for each sample. Both the 16p11.2-specific and genome-wide matrices were statistically analyzed to determine which factors decisively distinguish individuals with autism from those without. This analysis revealed that autism diagnosis is more commonly associated with 16p11.2 deletions than duplications or no 16p11.2 CNV. Analysis of the matrices also unveiled a key distinction: while autism diagnosis prediction using the 16p11.2 matrix is driven by the related developmental phenotypes, prediction using the genome-wide matrix is driven by the copy number of genes in the 16p11.2 region. It is our hope that this prototype model can be utilized in the clinic for additional phenotypes and disorders arising from genetic variants.

PrgmNr 2662 - CoverageMaster: a clinical grade and user oriented copy number variant caller

View session detail

Author Block: **M. Rapti**[1], M. Belfiore[2], E. Ranza[3], S. Antonarakis[4,3], F. Santoni[1,3]; [1]Ctr. Hosp.ier Univ.ire Vaudois (CHUV), Lausanne, Switzerland, [2]Genesupport, Ctr. of expertise in Genetics, Geneva, Switzerland, [3]Medigenome, Swiss Inst. of Genomic Med., Geneva, Switzerland, [4]Univ. of Geneva Med. Faculty, Geneva, Switzerland
Disclosure Block: **M. Rapti:** None.
Copy number variation (CNV) is the most frequent structural alteration in the human genome. Aberrant number of copies of specific genes or genomic regions (deletion: 0 or 1 copy, gain: >2 copies) are known to be implicated in pathogenic conditions such as Mendelian diseases and cancer. In clinical applications, while advances in next generation sequencing (NGS) have provided a standardized way for accurate coding variant analyses through whole exome sequencing (WES), CNV detection still relies on probe-based methods, such as array Comparative Genomic Hybridization (array CGH) and Multiplex Ligation-dependent Probe Amplification (MLPA). These technologies are complementary to NGS approaches in clinical diagnostics but limited array CGH resolution, poor MLPA scalability and additive costs are pushing for WES to become the primary strategy for identifying CNVs. However, WES technical issues such as lack of continuity of the target regions and biases due to the hybridization processes made it difficult so far to standardize a procedure for CNV detection. In particular, current WES based methods for CNV calling suffer from high false positive rates and reduced sensitivity for micro CNVs (

PrgmNr 2663 - Highly sensitive transcriptomic-based pooled CRISPR screening on 1 million cells enabled by spatial molecular imager

View session detail

Author Block: **S. Spisak**[1,2], M. L. Freedman[1,2,3], J-H. Seo[1,2], D. Kim[4], S. He[4], E. Piazza[4], P. Danaher[4], I. Lee[4], J. Jenkins[4], R. Liu[4], J. M. Beechem[4]; [1]Dana-Farber Cancer Inst., Boston, MA, [2]Ctr. for Functional Cancer Epigenetics, Dana-Farber Cancer Inst., Boston, MA, [3]The Eli and Edythe L. Broad Inst., Cambridge, MA, [4]NanoString Technologies, Seattle, WA
Disclosure Block: **S. Spisak:** None.
Pooled single-cell CRISPR-based screens present a powerful strategy to functionally link regulatory elements to their target genes. To date, most screens rely on readouts that can be robustly measured, such as proliferation and strong transcriptional responses. Measuring individual transcripts would be an ideal readout; however, accurate measurement of transcripts of interest at single cell resolution remains challenging.
Droplet and flow cytometry technologies coupled with next generation sequencing have been extended to detect mRNA and CRISPR guide (gRNA) sequences from the same cell for perturbation studies; however they tend to work optimally for high-expressing transcripts and for perturbations affecting pathways, limiting its broad applicability. Moreover, dropout rates impact efficiency and cost. Here, we describe a novel method using SMI (spatial molecular imaging), which precisely and quantitatively measures gRNAs and transcripts of interest within the same cell. This method is not restricted to polyadenylated transcripts and can accurately measure medium- and low-expressing transcripts at single cell resolution with throughput of hundreds of thousands to one million cells per experiment.
Methods: We performed pooled CRISPR-based screens by targeting an androgen receptor (AR) enhancer in the LNCaP prostate cancer cell line using SMI to investigate performance characteristics. LNCaP-modified lines were transduced by two gRNA pools (enhancer-targeting, positive/negative controls). AR and related target gene levels (N=37) were measured by SMI in a million cells and RT-PCR.
Results: Using SMI, we demonstrated quantitation of gene expression for AR and related target genes on over 90% of the cells analyzed (**Conclusion:** SMI enables simultaneous visualization of gRNA and accurate detection of low- and medium expressed transcripts at the single cell level. Looking forward, this platform has the potential to further improve resolution in pooled CRISPR screens by distinguishing transcript variants; analyzing multiple perturbation within single cells at throughput of hundreds of thousands of cells; and decreasing 3`-end transcript bias in gene expression readout.
Spatial molecular imager is for research use only and not for use in diagnostic procedures.

PrgmNr 2664 - Longitudinal multiomics data responses aid the detection of collective phenotypic characteristics in prediabetic/diabetic monitoring

View session detail

Author Block: **M. Zheng**[1], G. Mias[2]; [1]Biochemistry and Molecular Biology, IQ Ctr., Michigan State Univ., East Lansing, MI, [2]Michigan State Univ., East Lansing, MI
Disclosure Block: **M. Zheng:** None.
Longitudinal multiomics profiling, which combines data from biomolecular, physiological, environmental and clinical measures, presents great promise for precision health. Detecting temporal phenotypic characteristics across individuals by comparing the longitudinal multiomics is important for understanding and predicting disease outcomes based on molecular measures. A main challenge of comparing longitudinal multiple omics across individuals is data heterogeneity, including different scales, missing datapoints, uneven sampling, etc. We addressed this challenge by applying network-based analyses to categorize single individual's multi-omics, and use these to create multi-individual similarity clusters directly from their longitudinal multiomics profiles. These clusters revealed variability of phenotypic characteristics associated to immune responses in individualized multiomics from prediabetic and diabetic individuals.
We analyzed individualized multiomics profiles from public data for 69 individuals using spectral methods. We generated periodograms for individual subject omics signals, to construct within-person omics networks and analyze personal-level immune changes. We compared periodograms across all individuals to identify network clusters of individuals with similarities across their common omics temporal patterns.
Our analysis discovered both intra- and inter- individual characteristics differences. The inter-individual longitudinal omics clusters show pattern changes corresponding to the individualâ□□s physiological state changes. We identified similar individual-level responses to immune perturbation. The multi-individualsâ□□ similarity network revealed different classes within which the molecular behavior was linked to phenotypic differences, including body mass index and insulin resistance, with the immune response dominating differences attributed to diabetic status.
We detected collective phenotypic differences in prediabetic and diabetic individuals by creating intra- and inter- individual temporal multiomics networks. We captured common immune responses across diabetic and pre-diabetic individuals in response to immune perturbation (vaccination/infection). Based on collective temporal omics responses we identified individualsâ□□ clusters with differences in body mass index and insulin resistance. Our findings utilizing personal temporal omics to identify collective responses across individuals associated with macroscopic characteristics, can potentially help predict disease responses and outcomes towards clinical implementations.

PrgmNr 2665 - Meta-Learning on Next-Generation Sequencing data to Improve Cell Type Annotation

View session detail

Author Block: **Y. Park**, D. Heider, A-C. Hauschild; Philipps-Univ. of Marburg, Marburg, Germany
Disclosure Block: **Y. Park:** None.

Tremendous advances in next-generation sequencing technology during the last decade allowed the accumulation of vast amounts of omics data in various research areas. However, study limitations of small sample sizes, particularly in clinical research of rare diseases, technical heterogeneity, and batch effects, narrow the applicability of traditional statistics and machine learning analysis. Here, we present a meta-learning approach to transfer knowledge from large-scale data and reduce the search space in the data with small sample sizes. Few-shot learning algorithms integrate meta-learning to overcome data scarcity and data heterogeneity by transferring molecular pattern recognition models from datasets of unrelated domains. We investigate few-shot learning models with TCGA (The Cancer Genome Atlas) data and demonstrate its potential as a molecular pattern recognition model for small datasets. Our results show that transfer learning is very effective for datasets of limited sample sizes. Furthermore, we demonstrate that our approach can transfer knowledge across technological heterogeneity, e.g., bulk-cell to single-cell data. Our approach can overcome study size restrictions, batch effects, and technological limitations in single-cell data analysis by utilizing existing bulk-cell sequencing data.

PrgmNr 2667 - OUTRIDER2: A generalized framework for context-dependent outlier detection in omics data

View session detail

Author Block: **I. Scheller**[1], S. Loipfinger[2], D. Smirnov[3,4], R. Kopajtich[5], F. Brechtmann[2], H. Prokisch[6,3], J. Gagneur[7,1]; [1]Inst. of Computational Biology, Helmholtz Zentrum MÃ¼nchen, Neuherberg, Germany, [2]Faculty of Informatics, Technical Univ. of Munich, Garching, Germany, [3]Inst. of Human Genetics, Technical Univ. of Munich, Munich, Germany, [4]Inst. of Human Genetics, Helmholtz Zentrum Muenchen, Neuherberg, Germany, [5]Technical Univ. of Munich, Munich, Germany, [6]Helmholtz Zentrum Muenchen, Neuherberg, Germany, [7]Technical Univ. of Munich, Garching, Germany
Disclosure Block: **I. Scheller:** None.
The detection of aberrant gene expression events in RNA-seq data has recently been shown to be a promising complementary avenue to DNA sequencing both in rare disease diagnostics and in understanding effects of rare variants in common diseases. To this end, we have earlier adapted denoising autoencoders to RNA-seq data and implemented the method into the R package OUTRIDER [1]. OUTRIDER models expected RNA-seq gene read counts while controlling for latent confounders, and then performs significance based outlier detection using negative-binomial p-values. Here, we describe OUTRIDER2, an extension and generalization of OUTRIDERâ□□s denoising autoencoder and p-value based outlier detection method to continuous measurement and demonstrate it on mass-spectrometry based proteomics data. OUTRIDER2 adds support for modelling data with a Gaussian distribution and provides different data preprocessing and transformation options as well as a modularized model building procedure. It introduces a python backend and command line interface to the frontend R package which provides a flexible design that allows for an easy inclusion of additional distributions, data processing options and latent space fitting and decoding methods in the future. It furthermore supports the consideration of known covariates in addition to the data driven learning of the latent space during model fitting, and provides an improved runtime performance compared to the original R package for datasets with larger sample sizes. We demonstrate OUTRIDER2 by applying it to both transcriptomics and proteomics measurements within a rare disease cohort of 135 samples, enabling the joint analysis of aberrant RNA and protein expression events in the context of rare genetic variation [2]. Across all cutoffs, enrichments of rare variants suspected to disrupt protein abundance (missense, frameshift, stop gains) for outliers ranked by OUTRIDER2 is 100% higher than for rankings based on z-scores from regression on covariates (batch, sex) and 50% higher than for rankings p-values of differential expression tests (one-against-all, limma). In conclusion, this genome-wide observations highlight the complementarity of considering outlier events in different omics measurements to capture the functional impact of rare genetic variation, and showcases OUTRIDER2 as an easy-to-use tool that enables the study of aberrant events in a wider range of data modalities.
[1] Brechtmann, F. et al. OUTRIDER: a statistical method for detecting aberrantly expressed genes in RNA Sequencing Data. Am. J. Hum. Genet. (2018) [2] Kopajtich, R. et al. medRxiv (2021). doi: 10.1101/2021.03.09.21253187

PrgmNr 2668 - The value of long read genome sequencing in the diagnostic evaluation of muscular dystrophy

View session detail

Author Block: **C. C. Bruels**[1,2,3], E. S. McGaughy[3], E. A. Estrella[4,5], B. T. Darras[6,5], H. J. Stalker[7], C. D. Zingariello[3], Y. Wang[8], C. Collins[8], A. Behlmann[8], L. M. Kunkel[4,5], P. B. Kang[1,2,9,3]; [1]Paul and Sheila Wellstone Muscular Dystrophy Ctr., Univ. of Minnesota Med. Sch., Minneapolis, MN, [2]Dept. of Neurology, Univ. of Minnesota Med. Sch., Minneapolis, MN, [3]Div. of Pediatric Neurology, Dept. of Pediatrics, Univ. of Florida Coll. of Med., Gainesville, FL, [4]Div. of Genetics & Genomics, Boston Childrenâ□□s Hosp., Boston, MA, [5]Harvard Med. Sch., Boston, MA, [6]Dept. of Neurology, Boston Children's Hosp., Boston, MA, [7]Div. of Genetics, Dept. of Pediatrics, Univ. of Florida Coll. of Med., Gainesville, FL, [8]PerkinElmer Genomics, Atlanta, GA, [9]Inst. for Translational NeuroSci., Univ. of Minnesota Med. Sch., Minneapolis, MN

Disclosure Block: **C.C. Bruels:** None.

Introduction: Many patients with suspected muscular dystrophies remain genetically undiagnosed despite clinical diagnostic testing via second-generation short read sequencing and standard variant analyses of this sequence data. We propose that a significant proportion of these undiagnosed individuals harbor structural variants (SVs) that are not easily detected with this approach. Nanopore long-read whole genome sequencing, a form of third-generation sequencing, is a powerful technique that can be used to identify and resolve a higher proportion of such SVs, further identifying causal variants in these unsolved families. We optimized protocols to yield long reads from pre-existing DNA extractions and developed an analytical workflow to test our hypothesis. We detected a likely pathogenic SV that was initially undetected by short read sequencing in one family, and determined the relative positions of the copies of a multi-exon duplication in a second family.

Results: LAMA2-related dystrophies resulting from partial or complete merosin deficiency are one of the most common forms of recessive congenital muscular dystrophies. Clinical whole exome sequencing identified a single heterozygous pathogenic variant in *LAMA2* (c.2962C>T, p.Gln988X) in an individual with a sporadic congenital muscular dystrophy phenotype suggestive of merosin deficiency. Nanopore sequencing identified a 3,463 bp duplication (chr6:129660157-129663620) that includes all of exon 30, likely resulting in a frameshift and premature termination. We confirmed the presence of this duplication in 120-1 (proband) and 120-2 (mother) using PCR primers that span the boundaries of the duplicated region.

A second individual (1466-1) was found to have a duplication of exons 10-26 in *DMD* during a genetic evaluation for an unrelated issue. He is a 40-year-old construction worker with a normal CK and no weakness on physical examination, thus it was suspected that the duplication may be non-tandem. Nanopore sequencing was used to pinpoint the breakpoint locations, and determined that the duplication was indeed in tandem. This suggests that this SV may not be pathogenic for a Duchenne or Becker muscular dystrophy phenotype, but cardiac surveillance may be warranted. We confirmed the breakpoints of this duplication in 1466-1 using PCR and Sanger sequencing.

Conclusion: This study highlights the utility of nanopore sequencing to identify and characterize previously cryptic or incompletely defined SVs, potentially providing a more complete molecular diagnosis for patients, and increasing our understanding of the distribution of pathogenic variants in the setting of muscular dystrophy.

PrgmNr 2669 - Using deep learning to detect Noonan syndrome from electronic health records

View session detail

Author Block: **Z. Yang**, A. R. Shikany, G. Zhang, K. N. Weaver, J. Chen; Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH
Disclosure Block: **Z. Yang:** None.
Background: Noonan syndrome (NS) is a genetic disorder that affects an estimated 1 in 1000 to 2500 people. Common clinical features include short stature, heart defects, and congenital malformations. Variable expressivity is common making it difficult to obtain a timely diagnosis. Genetic testing can be used to confirm a diagnosis but many individuals do not get tested due to a lack of recognition and referral. Our study sought to investigate the utility of using electronic health records (EHR) to build deep learning models and to identify patients at high risk of NS.
Methods: Using diagnosis description text from Cincinnati Children's Hospitalâ□□s de-identified EHR database, we constructed deep learning models from 162 NS cases and 32400 presumable controls. Models were optimized using 6-fold cross-validation. The final models were tested with an independent test set, consisting of 28 NS cases and 28000 presumable controls.
Results: We found all deep learning models performed significantly better than our previously developed statistical method, Genetic Disease Diagnosis based on Phenotypes (GDDP). Convolutional neural net (CNN) had the best performance on NS classification among all techniques, with a 0.5 recall at 0.42 precision and 0.55 area under precision-recall curve.
Conclusion: The deep learning models based on diagnosis description text from EHR were able to classify NS patients and provide clinically relevant explanations. The results suggested the validity of this approach as a screening method for NS, and demonstrated the value of EHR data and machine learning methods for diagnostics of rare genetic diseases.

PrgmNr 2670 - Using transcriptomic analysis to identify pathogenic variants in rare Mendelian diseases missed on conventional genomic testing

View session detail

Author Block: **H. Safraou**[1,2], E. Colin[3,1], Y. Duffourd[4,2], A-S. DENOMME-PICHON[1,2], A-L. Bruel[5,2], V. Couturier[1,2], V. Bourgeois[1,2], M. Chevarin[1,2], C. Poe[1,2], L. Faivre[6,1,2], C. Philippe[1,2], C. Thauvin-Robinet[6,1,2], F. Tran Mau-Them[1,2], A. Vitobello[1,2]; [1]Inserm UMR1231 team GAD, Univ. of Bourgogne Franche-ComtÃ©, Dijon, France, [2]Functional Unity of innovative diagnosis for rare diseases, Dijon Bourgogne Univ. Hosp., Dijon, France, [3]Univ. Hosp. of Angers, Angers, France, [4] Inserm UMR1231 team GAD, Univ. of Bourgogne Franche-ComtÃ©, Dijon, France, [5]UMR1231 team GAD, Univ. of Bourgogne Franche-ComtÃ©, Dijon, France, [6]Genetics Ctr., FHU-TRANSLAD, Dijon Univ. Hosp., Dijon, France
Disclosure Block: **H. Safraou:** None.
The implementation of next-generation exome or genome sequencing (ES/GS) in clinical practice allows the rapid identification of causative genetic variants responsible for Mendelian diseases. However, a large part of patients with rare genetic disorders, and in particular with neurodevelopmental disorders (NDDs), remains without a molecular diagnosis, indicating that the genetic information alone is sometimes insufficient to identify causative variants. GS allows the analysis of regions not enriched by ES, and in particular intronic regions and poorly covered exons. Intronic, structural genomic variants and exonic variants not well covered by ES account for an additional diagnostic yield of about 10% as compared to ES. However, the challenge remains the clinical interpretation of the high amount of variants, especially those located within deep-intronic regions. Recently, it has been demonstrated that using of RNA sequencing (RS) as a complement of GS can improve diagnosis of patients with negative ES results. Indeed, RS can detect splicing anomalies that may be associated with aberrant gene expression (outliers).Here, we describe our experience about the implementation of RS in the laboratory practice for the analysis and the interpretation of transcriptome data in NDDs. We applied RS from whole blood and skin fibroblasts coupled with GS in a cohort of 100 patients with NDDs remained without molecular diagnosis despite standard genomics analysis (ES/GS). We implemented workflow to detect outliers in gene expression and splicing anomalies. Starting from sample handling to the generation of analysis report we describe potential caveats and pitfalls associated with RS. Our analyses identified cases with deep-intronic variants causing cryptic exon activation, expression down-regulation identified via transcriptomic sequencing and structural variants not previously identified by our GS pipeline.In conclusion, new diagnoses could be raised in our cohort of undiagnosed individuals with NDD and suspected rare Mendelian diseases using transcriptomic analysis.

PrgmNr 2671 - Clinical pharmacogenomic test reporting preferences among adult and pediatric providers

View session detail

Author Block: **A. Cera**[1], H. Naik[1], R. M. Wasserman[2], B. Duong[3], A. Owusu Obeng[1,4], K. J. Cook[5,6], N. Seligson[5,6], N. Cody[1,7], V. F. Funanage[3], K. V. Blake[5], S. A. Scott[1,8,9]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Nemours Children's Hlth., Orlando, FL, [3]Nemours Children's Hlth., Wilmington, DE, [4]Mount Sinai Hosp., New York, NY, [5]Nemours Children's Hlth., Jacksonville, FL, [6]Univ. of Florida Coll. of Pharmacy, Jacksonville, FL, [7]Sema4, Stamford, CT, [8]Stanford Univ., Stanford, CA, [9]Stanford Hlth.Care, Palo Alto, CA

Disclosure Block: **A. Cera:** None.

The increasing availability of practice guidelines has resulted in growing interest in implementing clinical pharmacogenomic testing; however, there is ongoing regulatory debate whether clinical pharmacogenomic test reports should include therapeutic recommendations. Previous studies have concluded that most providers do not feel equipped to independently interpret pharmacogenomic results. Therefore, we aimed to assess provider preferences on clinical pharmacogenomic report content. Adult (36%) and pediatric (60%) providers participated in an anonymous online national survey, including physicians (54%), pharmacists (22%), nurse practitioners (12%), and other providers (22%), which were diverse in age and gender, largely white (75%) or Asian (16%), and concentrated in urban and suburban medical centers in the Eastern U.S (n=105). Half of all respondents (54%) identified as being somewhat familiar with pharmacogenomics; however, only 36.5% had previously heard of the Clinical Pharmacogenetics Implementation Consortium (CPIC). Importantly, the majority (66%) of providers agreed that clinical pharmacogenomic reports should include identified genotypes, metabolizer phenotypes, potentially impacted medications and therapeutic recommendations, compared to only 7.7% who disagreed. Moreover, most providers agreed that all potentially impacted medications should be reported, including those with FDA label recommendations and from other sources (e.g., CPIC). Similarly, 72% of providers indicated that the therapeutic recommendations should include contraindicated medications, dosing information, and patient monitoring recommendations when applicable. Interestingly, physicians preferred more therapeutic recommendations and dosage guidance information than pharmacists (p=0.0007), suggesting that pharmacists may feel more confident than clinicians to translate pharmacogenomic results into therapeutic recommendations. Regarding panel testing, the majority (72%) of providers indicated they were not worried about including panel results in patient medical records. However, only ~30% of providers indicated they were not worried about their clinical responsibility for panel results beyond the original indication for the test or for panel results ordered by a different provider. Taken together, these data indicate that the majority of clinical providers prefer to have explicit therapeutic recommendations included in clinical pharmacogenomic test reports; however, additional resources are necessary to support the delivery of pharmacogenomic panel results to providers through electronic medical records.

PrgmNr 2672 - Common statin intolerance variants in *ABCB1* and *LILRB5* show synergistic effects in statin response: an observational study using electronic health records

View session detail

Author Block: **A. Melhem**[1], M. Chourasia[1], M. Bigossi[1,2], C. Maroteau[1], A. Taylor[1], R. Pola[2], A. Dawed[1], A. Tornio[1,3], C. N. Palmer[1], M. K. Siddiqui[1]; [1]Univ. of Dundee, Dundee, United Kingdom, [2]Fondazione Policlinico Univ.rio, Rome, Italy, [3]Univ. of Turku, Turku, Finland
Disclosure Block: **A. Melhem:** None.
Statin intolerance impacts approximately 10% of statin users, with side effects ranging from mild myalgia to extreme intolerance resulting in myopathy and rhabdomyolysis. Statin intolerance results in poor adherence to therapy and can impact statin efficacy. Many genetic variants are associated with statin intolerance. The effect of these variants on statin efficacy has not been systematically explored. Using longitudinal electronic health records and genetic biobank data from Tayside, Scotland, we examined the effect of nine genetic variants with previously reported associations with simvastatin or atorvastatin intolerance on the outcome of statin response. Statin response was measured by the reduction achieved when comparing pre-and post-statin non-HDL-cholesterol. Post-treatment statin response was limited to non-HDL-cholesterol measured within six months of therapy initiation. Univariate and multivariable linear regression models were used to assess the main and adjusted effect of the variants on statin efficacy. 9,401 statin users met study inclusion criteria, of whom 8,843 were first prescribed simvastatin or atorvastatin. The average difference in post-treatment compared to pre-treatment non-HDL-cholesterol was 1.45 (Â±1.04) mmol/L. In adjusted analyses, only two variants, one in the gene *ABCB1* (rs1045642) and one in *LILRB5* (rs12975366), were associated with statin efficacy. In *ABCB1*, homozygous carriers of the C allele at rs1045642 had 0.05 mmol/L better absolute reduction in non-HDL-cholesterol than carriers of the T allele (95%CI: 0.01,0.1). In *LILRB5* (rs12975366), carriers of the C allele had 0.05 mmol/L better absolute reduction compared to those homozygous for the T allele (95%CI: 0.01,0.08). When combined into a two-variant risk score, individuals with both the rs1045642-CC genotype and the rs12975366 -TC or CC genotype had a 0.05 mmol/L greater absolute reduction in non-HDL-cholesterol compared to those with rs1045642-TC or TT genotype and the rs12975366 -TT genotype (95%CI: 0.02, 0.08; P=0.002). We report two genetic variants for statin adverse drug reactions that are associated with statin efficacy. While the *ABCB1* variant has been shown to have an association with statin pharmacokinetics, no similar evidence for *LILRB5* has been reported. These findings highlight the value of genetic testing to deliver precision therapeutics to statin users.

PrgmNr 2673 - DrugnomeAI: An exome-wide machine-learning framework for druggability prediction and ranking of drug target candidates by stochastic semi-supervised learning

View session detail

Author Block: **A. Raies**, E. Tulodziecka, J. Stainer, A. R. Harper, S. Petrovski, D. Vitsios; AstraZeneca, Cambridge, United Kingdom
Disclosure Block: **A. Raies:** Salary/Employment; Arwa Raies is an employee of AstraZeneca.
Target selection is paramount to the overall success of clinical development activities. Targets with genomic support are more likely to succeed with large-scale genomic studies now yielding many putative targets. Target selection requires consideration for many aspects, including the druggability, or tractability, of a target. However, predicting druggability, using standard machine learning (ML) approaches, is challenging due to the small number of known targets, lack of robust negative data, and extreme data imbalance.To address these challenges, we developed drugnomeAI, an automated ML framework based on a previously published method (Vitsios & Petrovski, 2020) that has been re-purposed for estimating a druggability likelihood for every protein-coding gene in the human exome. DrugnomeAI implements a positive-unlabeled learning approach by randomly segregating the dataset into smaller balanced datasets of positive and unlabeled genes. DrugnomeAI provides specialised models stratified by both disease type (oncology and non-oncology) and therapeutic modality (small molecules or antibodies). The tool employs the Tclin annotation from PHAROS, as a labelled set of known drug targets, and integrates gene-level properties from 13 sources, including PHAROS, String, CTD and DGIdb as well as protein-protein interaction network data, resulting in a set of >350 features. The best performing model, based on gradient boosting, identifies known drug targets with 98% accuracy. Some key features contributing to the predictions include: protein-protein interactions from DGIdb, STRING, InWeb and Reactome, RNA specificity from Human Protein Atlas, monoclonal antibodies, as well as processes from CTD, such as apoptosis and phagosomes for the oncology-specific model.Investigation of the top ~1,000 drugnomeAI ranked genes shows that 74% have been previously selected for clinical development programs (48% in Phase IV). We found the top 5% of drugnomeAI gene target predictions to be significantly enriched for genes that achieved study-wide significance in a large PheWAS adopting 455K exomes from the UK Biobank. When compared to another ML tool (TargetDB), drugnomeAI demonstrates higher statistical enrichment in ranking genes from the Open Targets tractability data, both for small molecules (Fisherâ□□s exact: OR=4.1; $p=9.1x10^{-28}$) and antibodies (Fisherâ□□s exact: OR=3.7; $p=1.7x10^{-35}$).DrugnomeAI supports targets discovery by improving our understanding of the properties of druggable genes and prioritizing novel targets most amenable to small molecule and antibody modalities to further accelerate drug discovery.

PrgmNr 2674 - Pharmacogenetics for precision cardiovascular therapy: a novel GRS for *SLCO1B1* as a predictor of statin intolerance

View session detail

Author Block: **M. Bigossi**[1,2], C. Maroteau[3], A. Melhem[2], R. Pola[1], C. N. A. Palmer[2], M. K. Siddiqui[2]; [1]Section of Internal Med. and Thromboembolic Diseases, Dept. of Med., Fondazione Policlinico Univ.rio A. Gemelli IRCCS, L.go A. Gemelli 8, 00168 Rome, Italy, Rome, Italy, [2]Pat McPherson Ctr. for Pharmacogenetics & Pharmacogenomics, Div. of Population Hlth.& Genomics, Sch. of Med., Ninewells Hosp. & Med. Sch., Univ. of Dundee DD1 9SY, Dundee, United Kingdom, [3]Pat McPherson Ctr. for Pharmacogenetics & Pharmacogenomics, Div. of Population Hlth.& Genomics, Sch. of Med., Ninewells Hosp. & Med. Sch., Univ. of Dundee DD1 9SY, Dundee, Italy
Disclosure Block: **M. Bigossi:** None.
Statin intolerance impacts approximately 10% of statin users with side effects ranging from mild myalgia to myopathy and rhabdomyolysis. *SLCO1B1* has been extensively studied in relation to statin-associated muscle symptoms as it encodes OATP1B1, the main influx transporter of statins. However, studies have mostly focused on a single loss of function (LOF) variant (Val174Ala). Here we present findings that use several variants in *SLCO1B1* to produce a genetic risk score (GRS).
Using longitudinal electronic health records and genetic biobank data from Tayside, Scotland, we examined the effect of four common *SLCO1B1* variants with MAF > 0.05 on statin intolerance. Using *SLCO1B1* missense variants: Val174Ala, Asn130Asp, Pro155Thr, and Leu643Phe, an unweighted non-additive GRS for LOF was developed. The GRS was tested on two intolerance phenotypes: general statin intolerance (GSI), defined as having raised creatine kinase (CK) and 3 or more switches in statin therapy or discontinuation, and statin-induced myopathy/rhabdomyolysis (SIM) defined as statin users having CK levels â¥10 times the upper limit of normal in the ambulatory setting. These were compared to statin tolerant controls who had no recorded raised CK measures, were on therapy for a minimum of 5 years, had at least 90% daily coverage, had no more than 1 switch and no discontinuation of therapy. Logistic regressions were used to assess adjusted effect of the variants on statin intolerance phenotypes. Models were adjusted for statin type and dose, prior cardiovascular events, age and sex.
From 15,543 statin users, 521 individuals (3.4%) had GSI, while 108 (0.7%) had SIM. These were compared to 1901 statin tolerant individuals (12.2%). Individuals homozygous for Val174Val were classified as level 8 of the GRS. Risk alleles in other variants were treated additively with a maximum level of 7 (median 6, range 2-8). Individuals with GRS â¥7 (7% of study population) had significantly higher odds of GSI and SIM compared to those who had GRS â¤ 6. Individuals with GRS â¥7 had 2.42 times the odds of GSI compared to those with GRS â¤ 6 (95% CI 1.24-4.46, p 0.006). Similarly, individuals with GRS â¥7 had 3.50 times the odds of SIM compared to those with GRS â¤ 6 (95% CI 1.25-8.30, p 0.008). In contrast, Val174Ala alone is associated with a non-significant increase of both GSI (OR 1.94, 95% CI 0.64-5.88, p 0.24) and SIM (OR 2.76, 95% CI 0.64-12.0, p 0.18).
Impaired OATP1B1 function is associated with higher risk of statin intolerance. At present, Val174Ala in *SLCO1B1* is the only pharmacogenetic variant tested in clinical practice. This study shows that a comprehensive GRS of *SLCO1B1* is better able to predict risk of statin intolerance than Val174Ala.

PrgmNr 2675 - Sanford Health Preemptive Population Genetic Screening - A Comparison of Military Veterans and non-Veterans

View session detail

Author Block: **J. Leonhard**[1], A. Massmann[2], B. Tucker[3], J. Van Heukelom[2], K. Jacobsen[2], J. Kippley[2], A. Schultz[2], C. Hajek[2]; [1]Sanford Hlth., Bemidji, MN, [2]Sanford Hlth., Sioux Falls, SD, [3]Denver, CO
Disclosure Block: **J. Leonhard:** None.
Background: Population genetic screening programs for healthy adults are becoming more common even though guidelines regarding such testing are scant; however, considerations for implementation were recently published (1,2,3,4). Military veterans are a unique demographic of patients with complex healthcare needs (5). Mental health issues may be common, occurring in up to 33% of veterans (5). Many veterans take medications that have Clinical Pharmacogenetics Implementation Consortium guidelines (6,7). Efforts to implement pharmacogenomic (PGx) testing are underway in the Veterans Health Administration (6). In addition, up to 3% of veterans are expected to have secondary findings reported through the Sanford Health preemptive genetic screening program (Sanford Chip) (8). We present a comparison of results for military veterans and non-veterans from the Sanford Chip.
Methods: The Sanford Chip detects PGx and medically actionable predisposition (MAP) variants. Using a genotyping platform, we report the genotype/phenotype for up to 11 PGx genes as well as Sanger-confirmed pathogenic and likely pathogenic variants in 57 genes. Pharmacists review all PGx results, enter summaries in the electronic medical record (EMR), and notify providers of actionable results. Programmed decision support informs future medication or dose changes in the EMR. Patients with MAP-positive results are contacted and referred for clinical genetic counseling.
Results: During the period from 9/16/2020 through 5/15/2021, 437 veterans and 719 non-veterans had PGx results. All patients in this cohort had at least one PGx variant. Actionable drug-gene variants were present in 33.8% of veterans and 27.9% of non-veterans (p p = 0.21) were associated with an increased risk to develop disease, and the remainder were positive for carrier status or association with moderate disease risk.
Conclusion: Meeting the complex needs of veterans can be facilitated by implementing a personalized approach to their healthcare. Preliminary data presented here suggests preemptive genetic screening of veterans compared to non-veterans could impact medical management recommendations for veterans, especially with prescribing practices.

PrgmNr 2676 - Transcriptomic effects of propranolol and primidone uncover important pathways for tremor reduction in essential tremor

View session detail

Author Block: **C-E. Castonguay**[1], C. Liao[1], A. Khayachi[1], G. Houle[1], J. P. Ross[1], P. A. Dion[1], G. A. Rouleau[2]; [1]McGill Univ., Montreal, QC, Canada, [2]Montreal Neurological Inst.-Hosp., Montreal, QC, Canada

Disclosure Block: **C. Castonguay:** None.

Essential tremor (ET) is one of the most common movement disorders, with nearly 1% of the worldwide population suffering from the disease. Despite its high prevalence and heritability, few genetic risk factors for ET have been identified. Previously, studies of drug effects have been leveraged to uncover genes related to complex traits, particularly for disorders with subsets of drug-responsive patients. ET notably presents patients who are responsive to two tremorolytic drugs: propranolol and primidone. Genes targeted by these tremor reducing agents might be implicated in the pathophysiology of ET. The objective of this study was to identify convergent propranolol and primidone transcriptomic targets in cerebellar DAOY cells and cortical neural progenitor cells (NPCs) in order to inform on ET related genes. To achieve this, DAOY cells and NPCs were treated for 5 days with clinical concentrations of propranolol and primidone, followed by RNA-sequencing to identify differentially expressed genes (DEGs). A meta-analysis approach was used to identify convergent genes across the different cell lines and treatments. The expression of genes previously associated with ET, such as *TRAPPC11*, were significantly affected by treatment with propranolol. Furthermore, Reactome and Gene Ontology (GO) pathway enrichment analysis identified axon guidance, vascular endothelial growth factor (VEGF) signalling and endosomal sorting complex required for transport (ESCRT) as significant. Weighted gene co-expression network analysis (WGCNA) also revealed co-expression modules related to neurite and cell projection morphogenesis in NPCs as well as endosomal trafficking in DAOYs. Gene-set enrichment analysis in human cerebellar single-cell data also showed an enrichment of propranolol and primidone targeted genes in granule cells and endocytes. Loss-of-function analysis using observed/expected ratio (o/e) scores from gnomAD revealed that convergent primidone and propranolol DEGs had significantly higher mutational constraint than all protein coding genes, suggesting that these drugs affect the expression of genes that might carry rare deleterious variants. Our results therefore highlight important pathways by which propranolol and primidone might reduce tremor, as well as identify potential genes to test for rare variance in ET.

PrgmNr 2677 - An additionnal case of arthrogryposis with agenesis of the corpus callosum related to *SCYL2* diagnosed in the prenatal period

View session detail

Author Block: **A. JUVEN**[1,2,3], H. Safraou[1,2], G. Vera[4,5], E. Machevin[6], S. Patrier[7], M. Brasseur-Daubruy[8], F. Tran-Mau-Them[1,2], Y. Duffourd[9], L. Faivre[1,2,3], T. Christel[1,2,3], C. Philippe[1,2]; [1]UMR1231 GAD, Inserm - UniversitÃ© Bourgogne-Franche ComtÃ©, Dijon, France, [2]UnitÃ© Fonctionnelle Innovation en Diagnostic gÃ©nomique des maladies rares, FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon, France, [3]FÃ©dÃ©ration Hosp.o-Univ.ire MÃ©decine Translationnelle et Anomalies du DÃ©veloppement (TRANSLAD), Ctr. Hosp.ier Univ.ire Dijon, Dijon, France, [4]Dept. of Genetics and Reference Ctr. for Dev.al Disorders Normandy Ctr. and Personalized Med., Normandie Univ, UNIROUEN, Inserm U1245 and Rouen Univ. Hosp., Rouen, France, [5]Dept. of Pediatrics, Women and Child Dept., Ctr. Hosp.ier Eure Seine Evreux Hosp., Evreux, France, [6]Service de GynÃ©cologie-ObstÃ©trique, Ctr. Hosp.ier Eure-Seine, Evreux, France, [7]Dept. of Pathology, Rouen Univ. Hosp., Rouen, France, [8]Service de radiologie pÃ©diatrique et foetale, HÃ´pital Charles-Nicolle, Ctr. Hosp.ier Univ.ire de Rouen, Rouen, France, [9]CHU DIJON - INSERM U1231, DIJON, France

Disclosure Block: **A. Juven:** None.

SCYL2 is an evolutionary conserved pseudokinase that binds to clathrin and AP2, a plasma membrane adaptator. It phosphorylates the beta 2 adaptin subunit with a role in regulating excitatory receptor in the synapse. SCYL2 is important for neuronal survival in the developing brain. In mice, targeted disruption of *Scyl2* results in increased perinatal lethality, growth retardation and severe sensorimotor deficits (Gingras *et al. (*2015)). In humans, *SCYL2* bi-allelic nonsense and frameshift pathogenic variants were recently involved in arthrogryposis multiplex congenita 4, neurogenic, with agenesis of the corpus callosum (AMC4, OMIM : # 618766). Seidahmed et *al.(*2020) reported 8 individuals with AMC4 in 2 unrelated consanguineous families. Six of the 8 patients died in infancy or early childhood. Prenatal ultrasound included reduced fetal movements, polyhydramnios, clenched hands, fixed flexion of upper and lower limb joints, clubfeet, agenesis of the corpus callosum and dilated brain ventricles. At birth, the patients presented with arthrogryposis, microcephaly, facial dysmorphology, shortening of the femurs, fractures of the long bones. Most patients had seizures and peripheral neuropathy. We describe here an additionnal case of a male fÅtus from unrelated parents with AMC4. Ultrasonography at 24 weeks of amenorhea showed reduced fetal movements, corpus callosum agenesis, microretrognathia, clubfeet and clenched hands. No chromosomal abnormalities were detected by array CGH. Prenatal trio exome sequencing showed the fÅtus carried two compound heterozygous pathogenic variants in *SCYL2* (NM_001317784.1:c.97delG, p.Asp33Metfs* and c.176dup, p.Glu60Glyfs*8). None of these variants are found in a homozygous state in the GnomAD database. Ultrasonographic signs in this fÅtus are compatible with the diagnosis of AMC4 confirming these variants are implicated in the phenotype. Our study reports the first case of AMC4 diagnosed in the prenatal period and gives additionnal support for the involvement of *SCYL2* in autosomal recessive arthrogryposis in humans.

PrgmNr 2678 - Association between acute maternal infections during pregnancy and birth defects in Bogota and Cali, Colombia 2001-2018

View session detail

Author Block: **J. Rumbo**[1], I. Madariaga[2], A. Salazar[1], M. Sierra[1], C. Tovar[2], D. Sequera[1], C. Rodriguez[3], P. Hurtado[4], K. Sarmiento[5], I. Zarante[6]; [1]Pontifical Xavierian Univ., Bogota, Colombia, [2]Inst. of Human Genetics. Pontifical Xavierian Univ., Bogota, Colombia, [3]SecretarÃa Distrital de Salud de BogotÃi, Bogota, Colombia, [4]Pontifical Xavierian Univ., Cali, Colombia, [5]Dept. of Physiological Sci.. Faculty of Med.. Pontifical Xavierian Univ., Bogota, Colombia, [6]Genetics Service, Hosp. Univ.rio San Ignacio, Bogota, Colombia

Disclosure Block: **J. Rumbo:** None.

Infectious diseases occur frequently during pregnancy and their teratogenic role is well established in TORCH (Toxoplasmosis, other infections, Rubella, Cytomegalovirus, and Herpes simplex). However, the intrauterine development consequences of the rest of the infectious diseases that affect gravid women are unknown. We described a cohort of patients with major congenital defects (CD) and the exposure to infectious diseases during pregnancy using the information of Congenital Defects Surveillance Programs of Bogota and Cali (Colombia) between 2001 and 2018. We evaluated associations between groups of maternal infectious diseases and CD among 3,096 cases and 7,446 controls. CD presentation was more frequent as isolated (64.3%), polymalformed (23.2%), and syndromic (12.4%). 52.5% of cases and 44.6% of controls presented infectious diseases during pregnancy. The most common polyinfection was vaginal and urinary tract infection. The most common single infection was vaginal infection between cases and controls. We found an association between CD and gastrointestinal infections with an odds ratio (OR) 2.06 (IC 1.18 - 3.59), urinary tract infections OR 1.16 (IC 1.05 - 1.30), viral infections OR 1.88 (IC 1.18 - 3.0), respiratory infections OR 1.56 (IC 1.28 - 1.9), and vaginal infections OR 1.18 (IC 1.08 - 1.30). Comprehending the teratogenic effect of infectious diseases is relevant to promote prevention, screening, timely diagnosis, and appropriate treatment to gravid women.

PrgmNr 2679 - Evidence for genetic imprinting and a causal role of fetal insulin and diastolic blood pressure in regulation of placental weight

View session detail

Author Block: **R. N. Beaumont**[1], C. Flatley[2], M. Vaudel[3], G-H. Moen[4], L. Wegner ThÃ¸rner[5], H. Ullum[5], H. Svarre Nielsen[6], S. Brunak[6], J. Fadista[7], R. Fore[8], K. Trajanoska[9], J. Felix[9], C. SÃinchez-Soriano[10], S. Srinvasan[11], S. E. Stinson[6], E. Nohr[12], C. J. Willer[13], B. Ã svold[14], J. Bilbao[15], M. Bustamante[16], J. Ronkainen[17], C. Wang[18], C. E. Pennell[18], M. Atalay[19], T. Lakka[19], M-F. Hivert[20], M-R. Jarvelin[21], B. Jacobsson[22], B. Feenstra[23], P. R. Njolstad[24], Early Growth Genetics (EGG) Consortium; [1]Univ. of Exeter Med. Sch., Exeter, United Kingdom, [2]Norwegian Inst. of Publ. Hlth., Oslo, Norway, [3]Univ Bergen, Bergen, Norway, [4]Oslo Univ. Hosp., Oslo, Oslo, Norway, [5]Copenhagen Univ. Hosp., Copenhagen, Denmark, [6]Univ. of Copenhagen, Copenhagen, Denmark, [7]Statens Serum Inst., Copenhagen, Copenhagen, Denmark, [8]Harvard Pilgrim Hlth.care Inst., Boston, MA, [9]Erasmus Med. Ctr., Rotterdam, Netherlands, [10]Univ. of Edinburgh, Edinburgh, United Kingdom, [11]Univ. of Dundee, Dundee, United Kingdom, [12]Univ. of Southern Denmark, Odense, Denmark, [13]Univ. of Michigan, Ann Arbor, MI, [14]Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [15]Univ. of the Basque Country, Leioa, Spain, [16]ISIGlobal, Barcelona, Spain, [17]Univ. of Oulu, Oulu, Finland, [18]Univ. of Newcastle, Newcastle, Australia, [19]Univ. of Eastern Finland, Kuopio, Finland, [20]Harvard Pilgrim Hlth.Care Inst., Boston, MA, [21]Imperial Col London, London, London, United Kingdom, [22]Sahlgrenska Univ Hosp, Gothenburg, Sweden, [23]Statens Serum Inst., Copenhagen, Denmark, [24]Haukeland Univ Hosp., Bergen, Norway

Disclosure Block: **R.N. Beaumont:** None.

Introduction The placenta plays an essential role in fetal development. Placental weight can be indicative of efficacy; lowered by poor placentation, or increased under excessive gestational weight gain or diabetes. Very high or low placental weights associate with adverse outcomes such as preeclampsia, reduced fetal growth, and preterm birth. Despite this, little is known about genetic underpinnings of human placental growth. **Methods** We meta-analysed genome-wide association statistics on placental weight adjusted for sex and gestational age at 11,217,023 SNPs in 65,651 offspring, 60,886 mothers and 52,400 fathers of predominantly European ancestries in 19 cohorts. We used these results in Mendelian Randomization analyses to examine the causality of maternal exposures on placental weight. **Results** We identified 32 fetal, 4 maternal and 2 paternal loci at genome-wide significance (p-8). One paternal locus was distinct from the nearby fetal locus, the other colocalised with a fetal signal. Of the 37 independent loci, 17 are either novel or distinct from nearby birth weight SNPs. We estimated independent parental and fetal effects at the loci using a novel weighted linear model, providing evidence that the maternally-, but not paternally-inherited C allele at rs2237892 (KCNQ1 fetal locus) was associated with lower placental weight. Mendelian Randomization showed that higher maternal glucose and lower maternal insulin secretion causally increase placental weight, with opposite effects using fetal SNPs. Using blood pressure SNPs, we found evidence that higher fetal, but not maternal, genetically instrumented diastolic blood pressure lowers placental weight. **Conclusion** We identified 37 loci associated with placental weight, the majority having only direct fetal effects. Substantial overlap exists with loci previously associated with birth weight, with evidence of imprinting at locus KCNQ1. Imprinting effects at this locus have previously been shown for Type 2 Diabetes (T2D risk allele lowers placental weight). Mendelian Randomization supported the role of fetal insulin in regulation of placental growth and, in contrast to what we previously observed for birth weight, fetal but not maternal genetically predicted diastolic blood pressure potential alters placental weight.

PrgmNr 2680 - Genome-wide genetic control of fetal placental genomics shows multiple associations with health and disease across the life course, informing the Developmental Origins of Health and Disease

View session detail

Author Block: **A. Bhattacharya**[1], A. N. Freedman[2], V. Avula[2], R. Harris[2], W. Liu[2], C. Pan[1], A. Lusis[3], R. M. Joseph[4], L. Smeester[2], H. J. Hartwell[2], K. C. K. Kuban[5], C. Marsit[6], Y. Li[2], T. O'Shea[2], R. C. Fry[2], H. P. Santos, Jr.[2]; [1]Univ. of California at Los Angeles, Los Angeles, CA, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Univ. of California, Los Angeles, Los Angeles, CA, [4]Boston Univ., Boston, MA, [5]Boston Univ. Sch. of Med., Boston, MA, [6]Emory Univ., Atlanta, GA
Disclosure Block: **A. Bhattacharya:** None.
The placenta is the master regulator of the intrauterine environment and is central to the Developmental Origins of Health and Disease (DOHaD). Studies show that fetal genetics and placental genomics can influence child health traits. An integrative analysis of genetics, placental genomics, and child health traits has not been done and would yield insight into the DOHaD hypothesis. Recently, we developed Multi-Omic Strategies for Transcriptome-Wide Association Studies (MOSTWAS), which uses mediation analysis to scan variants genome-wide, detect gene-trait associations (GTAs), and develop hypotheses for trait-associated gene regulation. Here, using genetic, transcriptomic, and methylomic data from the Extremely Low Gestational Age Newborn (ELGAN) Cohort Study (N = 272), we applied MOSTWAS to train genetic models of expression of all genes on the ELGAN RNA-seq panel, 2,994 of which showed strong in- and out-sample accuracy. With these models, we conducted TWAS for 40 traits from 5 categories (autoimmune, metabolic, cardiovascular, perinatal, neuropsychiatric) and identified 264 GTAs across 176 TWAS genes and potential transcription (TFs) or epigenomic factors regulating their expression. Of the 176 genes, 50 were associated with multiple traits, many not genetically correlated (e.g., *IDI1* with BMI and schizophrenia). Genetically-regulated placental expression, in aggregate, explained significant portions of three neonatal traits (total puberty growth, childhood BMI, start of puberty) at 5-8% of total SNP heritability. In addition, 91 GTAs showed significant associations through distal variants, with many mediated through 8 TFs associated with multiple TWAS genes. For example, *EPS15*, a maternally imprinted placental TF associated with fetal growth, showed a negative association with waist-hip ratio (WHR) and was negatively associated with two genes: *SPATA13* and *FAM214A*, both showing positive TWAS associations with WHR. In human placenta-derived trophoblasts, FANA silencing of *EPS15* led to upregulation of both *SPATA13* and *FAM214A*, providing evidence for placental TF regulation of these genes. Further transcriptomic and functional consequences of *EPS15* knockdown are under evaluation. Our study reveals potentially shared placental pathways associated with many traits across the life course. GTAs with traits across categories and different life periods suggest that placental dysregulation affects fundamental early-in-life traits, and these effects compound and manifest in later-in-life traits. Our work motivates increased sample sizes for early childhood trait GWAS and the placenta as a key tissue of study.

PrgmNr 2681 - Investigating the influence of fetal growth on maternal blood pressure in pregnancy

View session detail

Author Block: **C. Decina**[1,2,3], R. N. Beaumont[1], N. M. Warrington[2,3], K. A. Patel[1], W. L. Lowe[4], D. M. Evans[2,3,5], R. M. Freathy[1,5]; [1]Inst. of BioMed. and Clinical Sci., Coll. of Med. and Hlth., Univ. of Exeter, Exeter, United Kingdom, [2]Inst. for Molecular BioSci., Univ. of Queensland, Brisbane, Australia, [3]Univ. of Queensland Diamantina Inst., Univ. of Queensland, Brisbane, Australia, [4]NU Feinberg Sch Med, Chicago, IL, [5]Med. Res. Council Integrative Epidemiology Unit, Univ. of Bristol, Bristol, United Kingdom
Disclosure Block: **C. Decina:** None.
Raised maternal blood pressure (BP) confers a higher risk of pregnancy complications such as preeclampsia and fetal growth restriction. The role of fetal growth variationâ□□s contribution to maternal BP variation is not well understood. Prior evidence includes (i) cases of Beckwith-Wiedemann syndrome (fetal abnormalities of chr11p15) associated with fetal overgrowth and maternal gestational hypertension and (ii) recently published work demonstrating that fetal birth weight (BW)-increasing alleles were associated with higher maternal BP. We aimed to further explore fetal influence on maternal BP in pregnancy by (1) testing its association with fetal sex given the known difference in BW (males ~130g heavier than females) and (2) testing its association with a fetal genetic score (GS) for BW in further samples.

We performed multivariable linear regression to explore the association between fetal sex and maternal systolic and diastolic BP (SBP & DBP) measured at 28 wk gestation from term, singleton births, in women from the HAPO (n=12,185) and ALSPAC (n=11,875) studies. We adjusted models for maternal age, parity, gestational age (GA), and a BW z-score either adjusted for sex and GA at delivery or unadjusted. We meta-analysed effects across cohorts. We also built a fetal BW GS using 190 BW SNPs adjusted for maternal effects and meta-analysed its association with maternal SBP and DBP in HAPO (n=915) and ALSPAC (n=7,499) mother-child pairs of European-ancestry.

Male fetal sex was associated with a 2.9 mmHg/kg higher maternal SBP [95%CI: 1.2 - 4.7], the 95%CI including the estimated coefficient for BW of 1.9 mmHg/kg higher SBP [95%CI: 1.7 - 2.2] in the adjusted BW model, with no sex association with SBP in the unadjusted BW model (P=0.31), consistent with mediation of a sex effect by BW. Male fetal sex was associated with a 2.9 mmHg/kg higher DBP [95%CI: 1.7 - 4.2], similar to a BW effect of 3.7 mmHg/kg higher DBP [95%CI: 3.1 - 4.3] in the adjusted model, with a possible small independent sex effect seen in the unadjusted model (P=0.005). Our fetal BW GS analyses found a 0.1 mmHg higher maternal SBP [95%CI: 0.0 - 0.3] and DBP [95%CI: 0.0 - 0.2] per 1 SD higher GS.

Maternal BP contributes to pregnancy complication risk, thus understanding whether fetal sex is a contributor alongside other risk factors could inform clinical care. Results found here suggest that male sex is associated with higher maternal SBP and DBP, and that these effects are likely driven by larger male size. Evidence of a sex effect on DBP independent of BW requires further investigation. We found fetal alleles that increase BW show weak evidence of association with higher maternal BP, but larger samples are needed to confirm this.

PrgmNr 2682 - MD

View session detail

Author Block: **S. JuÃirez-PeÃ±alva**[1,2,3], L. Schottlaender[4,1,5], M. Vallone[2], F. Cardoso[1], A. Real[1], M. Romano[1], P. Brenzoni[1], G. Musante[1,6], S. Sevilla[1], M. RodrÃguez Zubieta[1,7], A. Etchegaray[3], K. Tuminelli[8], L. B. Gordon[9,10,8]; [1]Inst. de medicina genÃ³mica (IMeG), Hosp. Univ.rio Austral, Univ. Austral, Pilar, Argentina, [2]Servicio de genÃ©tica mÃ©dica, Hosp. Univ.rio Austral, Univ. Austral, Pilar, Argentina, [3]Div. of Fetal Med., Hosp. Univ.rio Austral, Pilar, Argentina, [4]Inst. de Investigaciones en Med. Traslacional (IIMT), CONICET-Univ. Austral, Pilar, Argentina, [5]Servicio de NeurologÃa, Hosp. Univ.rio Austral, Univ. Austral, Pilar, Argentina, [6]Servicio de neonatologÃa, Hosp. Univ.rio Austral, Univ. Austral, Pilar, Argentina, [7]Servicio de AnatomÃa PatolÃ³gica, Hosp. Univ.rio Austral, Univ. Austral, Pilar, Argentina, [8]The Progeria Res. Fndn., Peabody, MA, [9]Dept. of Pediatrics, Div. of Genetics, Hasbro Childrenâ□□s Hosp. and Warren Alpert Med. Sch. of Brown Univ., Providence, RI, [10]Dept. of Anesthesiology, Perioperative and Pain Med., Boston Childrenâ□□s Hosp. and Harvard Med. Sch., Boston, MA

Disclosure Block: **S. JuÃirez-PeÃ±alva:** None.

Progeroid syndromes are rare, monogenic disorders that present with clinical features resembling accelerated aging. Hutchinson-Gilford progeria syndrome (HGPS) is the most common form, typically developing in childhood and causing premature death within the first two decades of life. Most HGPS cases are autosomal dominant, caused by a single *de novo* synonymous change causing altered splicing in exon 11 of the *LMNA* gene. *LMNA* encodes for lamin A, an essential protein for nuclear structure and function.

We report a male patient presenting with intrauterine growth retardation, hand contractures, microcephaly and microretrognathia. There is no family history of neurological problems; parents are Caucasian and unrelated. The patient presented with low birth weight (##1800g), erythroderma, thin and transparent skin, lagophthalmos, sparse scalp hair and mandibular underdevelopment. Hematological findings were high platelet count and anemia. Perinatally, there were difficulties in weight gain with poor suction during breastfeeding, precipitating nasoenteric feeding. At age 6 months, the patient continued to require nasoenteric feeding.

Genomic investigations included a normal karyotype and normal array-CGH. Whole exome sequencing revealed a heterozygous missense variant in exon 6 of the *LMNA* gene. The variant, NM_170707.2:c.973G>A,p.(Asp325Asn), was not present in publicly available databases (gnomad, exome variant server, 1000 genomes), was well conserved across species and predicted to be pathogenic *in silico*. According to the ASHG variant classification guidelines (2015), it was catalogued as a variant of unknown significance. However, The Progeria Research Foundation (International Progeria Registry and Cell and Tissue Bank) has identified an as yet unreported previous case with the same variant and neonatal presentation, whereby the patient had similar clinical features and passed away at age 9 months.

In summary, we describe an interesting patient presenting in fetal life with a progeroid syndrome associated with a novel *LMNA* variant. Early diagnosis and molecular characterization of such unusual diseases are essential for genetic counselling, the participation of patients in clinical trials, and the development of biomarkers and targeted therapies. Current and future substantial advances in diagnosis and management of progeria are only possible through international clinical and scientific collaborations.

PrgmNr 2683 - Mining NIPT data of 108.349 Dutch pregnant women: cfDNA based viral analysis and GWAS

View session detail

Author Block: **E. Sistermans**, J. Linthorst; Amsterdam UMC, Amsterdam, Netherlands
Disclosure Block: **E. Sistermans:** None.
Objectives: Non Invasive Prenatal Testing aims at detecting fetal aneuploidies using maternal plasma. The sequencing data generated for NIPT can be repurposed for other goals, such as the detection of viral DNA and the imputation of SNPs, thereby allowing GWAS. As viral infections during pregnancy are a major health concern to both mother and fetus, our first goal was to characterize the DNA virome in a cohort of 108.349 pregnant women. These data will enable future implementations for the early detection of specific viral infections during pregnancy. Next, we performed different genome-wide association studies. We confirmed our method by performing a GWAS targeted at length and BMI. Consequently, we looked for novel loci that associate with factors that affect NIPT performance such as the fraction of fetal cfDNA and the overall concentration of cfDNA. Finally, we combined the viral and GWAS studies to discover genes that associate with the previously detected viral infections. **Results**: In our anonymized cohort, we identified the presence of several viral species that are potentially harmful to mother or fetus or both, including Herpesvirus 5 (Cytomegalovirus), Parvovirus B19 and Hepatitis B. For each viral species detected we report an estimate of the prevalence in our Dutch cohort. Most viruses were detected at a low abundance, but we also observed cases with an exceptionally high viral load of Parvovirus B19 and Hepatitis B. DNA fragments originating from different viruses were found to have distinct fragmentation patterns, presumably caused by nucleosome remodeling, topology and timing of the infection. The GWAS analysis targeted at BMI and height replicated many known loci, while our GWAS into factors that affect NIPT performance highlighted genes involved in DNA degradation and autoimmune disease. Finally, by combining the most commonly detected viral infections with the imputed variants, we identified genes that potentially drive the susceptibility to viral infections including specific viral receptor genes and the HLA locus. **Conclusion**: We demonstrate that with a straightforward approach, we are able to detect viral DNA from typical genome-wide NIPT cfDNA sequencing data and to report the main characteristics of the virome in our cohort. We also demonstrate that SNPs imputed from low coverage NIPT sequencing data can be used for GWAS analyses. Future efforts are directed at elucidating genes that are involved in pregnancy associated problems such as pre-eclampsia and preterm birth.

PrgmNr 2684 - Small supernumerary marker chromosome 15 in Prenatal Diagnosis: A clinical challenge

View session detail

Author Block: **J. Catanho da Silva**[1], S. Lemos Ferreira[2], M. Santos Melo[2], T. Kay[1], M. VenÃ¢ncio[1], I. Carvalho[1]; [1]Dept. of Med. Genetics, Hosp. Dona EstefÃ¢nia, Centro Hosp.ar UniversitÃirio Lisboa Central, Lisbon, Portugal, [2]

Disclosure Block: **J. Catanho da Silva:** None.

Background: Small supernumerary marker chromosomes (sSMCs) are rare cytogenetic abnormalities, present in less than 0.08% of all pregnancies. sSMCs are extra chromosomes with structural anomalies whose size is smaller than or equal to that of chromosome 20 on the same metaphase spreads. Reportedly, about 70% of cases of sSMCs are *de novo* and most frequently derived from chromosome 15. Uniparental disomy (UPD) is the inheritance of both homologous chromosomes from the same parent. The most prominent clinical outcome of UPDs are imprinting disorders, and when affecting chromosome 15, is associated with Prader Willi or Angelman syndromes.

Case Presentation: We report a 39-year-old woman that underwent amniocentesis at 16 weeks of gestation, due to high risk for trisomy 21 (1/88) at first trimester combined screening. Ultrasound performed at 21 weeks and 5 days of gestation showed bilateral clubfoot. Conventional cytogenetics identified an abnormal fetal karyotype: 47,XY,+mar[7]/46,XY[10]. Parents had normal karyotypes. Fetal chromosomal microarray (CMA) analysis was normal. Fluorescence *in situ* hybridization (FISH) revealed a marker chromosome del(15)(q10), consisted exclusively of heterochromatic material. UDP studies indicated maternal UDP of chromosome 15, establishing the diagnosis of Prader-Willi syndrome. After genetic counseling, parents decided on the termination of pregnancy.

Discussion: Identification of sSMC is a difficult task in prenatal diagnosis (PND). Conventional karyotype analysis can detect numerical and structural chromosomal abnormalities but cannot determine the origin and genetic content of sSMCs. The identification of a sSMC in prenatal settings must alert clinicians to a possible UPD. Prenatal UPD testing should be considered, among other criteria, when a structurally abnormal chromosome 15 or a *de novo* sSMC with no apparent euchromatic material in the fetus are identified.

Conclusion: We bring to attention the importance of combining cytogenetic and molecular genetic techniques to the detailed characterization of marker chromosomes in PND. It allows proper genetic counseling, facilitates informed decision-making, prevents uncertainty, and provides proper prognostic assessments.

PrgmNr 2685 - A computationally efficient clustering linear combination approach to jointly analyze multiple phenotypes for GWAS

View session detail

Author Block: **M. Wang**, S. Zhang, Q. Sha; Mathematical Sci., Michigan Technological Univ., Houghton, MI
Disclosure Block: **M. Wang:** None.
There has been an increasing interest in joint analysis of multiple phenotypes in genome-wide association studies (GWAS) because jointly analyzing multiple phenotypes may increase statistical power to detect genetic variants associated with complex diseases or traits. Recently, many statistical methods have been developed for joint analysis of multiple phenotypes in genetic association studies, including the Clustering Linear Combination (CLC) method. The CLC method works particularly well with phenotypes that have natural groupings. In the CLC method, one first clusters individual statistics from the association tests for each phenotype into positively correlated clusters, then the CLC test statistic was developed to combine the individual statistics linearly within each cluster and combine the between-cluster terms in a quadratic form. It was theoretically proved that if the individual statistics can be clustered correctly, the CLC test statistic is the most powerful test among all tests with certain quadratic forms. Due to the unknown number of clusters for a given data, the final test statistic of CLC method is the minimum p-value among all p-values of the CLC test statistics obtained from each possible number of clusters. Therefore, a simulation procedure must be used to evaluate the p-value of the final test statistic. This makes CLC method computationally demanding. We develop a new method called computationally efficient CLC (ceCLC) to test the association between multiple phenotypes and a genetic variant. Instead of using the minimum p-value as the test statistics in CLC, ceCLC uses the Cauchy combination test to combine all p-values of the CLC test statistics obtained from each possible number of clusters. The test statistic of ceCLC follow a standard Cauchy distribution, so the p-value can be obtained from the cumulative density function without the need of the simulation procedure. Through extensive simulation studies and application on the COPDGene data, the results demonstrate that the type I error rates of ceCLC are effectively controlled in different simulation settings and ceCLC either outperforms all other methods or has statistical power that is very close to the most powerful method with which it has been compared.

PrgmNr 2686 - A method for estimating personalized causal genomic variants of disease

View session detail

Author Block: **M. A. Rahman**[1], C. Cai[2], N. Bo[3], D. McNamara[4], Y. Ding[3], G. F. Cooper[2], X. Lu[2], J. Liu[1,5]; [1]Dept. of Engineering Management and Systems Engineering, Missouri Univ. of Sci. and Technology, Rolla, MO, [2]Dept. of BioMed. Informatics, Univ. of Pittsburgh, Pittsburgh, PA, [3]Dept. of Biostatistics, Univ. of Pittsburgh, Pittsburgh, PA, [4]Dept. of Med., Univ. of Pittsburgh, Pittsburgh, PA, [5]Dept. of Biological Sci., Missouri Univ. of Sci. and Technology, Rolla, MO
Disclosure Block: **M.A. Rahman:** None.
Genomic variants of disease are often discovered nowadays through population-based genome-wide association studies (GWAS). These genomic variants are not necessarily present in each individual with the disease due to inter-individual genomic heterogeneity. Such individual-level genomic variation is important for designing personalized treatment, which is, however, not generally captured well by population-level models, such as GWAS. In addition, GWAS typically requires a large sample to detect low-frequency genomic variants with sufficient power. To help address these challenges, we have developed an instance-specific causal inference (ICI) algorithm for estimating the causal genomic variants that influence traits at the level of an individual (e.g., patient). In particular, ICI is a Bayesian method for learning instance-specific causal Bayesian networks from observational data; to our knowledge, it is unique in doing so. By modeling at the level of an individual, ICI is capable of finding low-frequency or rare genomic variants that may exist only in a small number of individuals. We compared the ability of ICI to find causal genomic variants of clinical phenotypes, relative to GWAS. In particular, from the Framingham Study (FHS; dbGaP Study Accession: phs000007.v30.p11) we used whole genome data from Affymetrix HuGeneFocused50K measurements and harmonized phenome data of blood pressure measurements and plasma lipid levels. ICI and GWAS identified a significant number of shared variants; the unique variants identified only by ICI have a much lower minor allele frequency (MAF) than those identified by GWAS. In addition, we observed that ICI discovered more individualized and diverse variants with higher likelihood ratios among the hypertension patients than did GWAS. Also, ICI identified top-ranked causal variants that predicted hypertension better than did GWAS, according to the area under the ROC curve. Finally, ICI found several well-known low-frequency causal variants of serum lipid levels that were missed by GWAS. Overall, the results provide support for ICI as a promising approach for learning personalized causal genomic variants of clinical traits and disease to help advance precision medicine.

PrgmNr 2687 - A powerful test of ancestral heterogeneity in the effects of gene expression on complex traits

View session detail

Author Block: **K. Knutson**[1], W. Pan[2]; [1]Div. of Biostatistics, Univ Minnesota, Minneapolis, MN, [2]Univ Minnesota, Minneapolis, MN
Disclosure Block: **K. Knutson:** None.
The Transcriptome Wide Association Study (TWAS) is a widely used approach which integrates expression and GWAS data to study the role of cis-regulated gene expression (GEx) in complex traits. TWAS models GEx as a function of cis-eQTL genotypes. However, strong evidence suggests that the genetic architecture of GEx varies across populations. Furthermore, recent findings point to possible ancestral heterogeneity in the effects of GEx on complex traits, heterogeneity which may be amplified in TWAS by modeling GEx as a function of cis-eQTLs. We present a novel extension to TWAS which models heterogeneity in the effects of cis-regulated GEx which are correlated with ancestry. By jointly analyzing samples from multiple populations, our multi-ancestry TWAS framework can improve power to detect genes with shared expression-trait associations across populations through increased sample sizes, as compared to existing stratified TWAS approaches. Under our proposed model, we derive score tests for homogeneous, heterogeneous, and total GEx effects on a complex trait. Our preliminary simulations reveal conserved Type-I error rates and high power across a number of scenarios, holding promise for further simulations on larger simulated datasets. We apply our test to case-control genotypes from the Alzheimer's Disease Sequencing Project (ADSP) and publicly available prediction models from the Multi-Ethnic Study of Atherosclerosis (MESA) study. We identify a number of genes with suggestive heterogeneous effects between African American and Caucasian subjects on Alzheimerâ□□s Disease (AD), including *ASPHD2*, *SDSL*, *ZNF589*, *PGM2L1*, and *PPIL3*. Our preliminary application further identifies many putative AD risk genes which were not discovered through ancestry-stratified TWAS analyses, many of which have well established or biological plausible links to AD, including (but not limited to) *RTP4*, *ORAI2*, and *TOMM40L*. In forthcoming work, we will apply our test to a larger sample from ADSP, anticipating a greater number of significant findings due to the increased sample size. Additionally, we will apply our proposed test of heterogeneity to continuous endophenotypes from the UK Biobank (n â□□ 500,000), specifically considering a set of imaging derived phenotypes with strong associations to AD.

PrgmNr 2688 - An ensemble tool to predict health outcomes using microbiome data

View session detail

Author Block: **T. Wang**[1], B. Li[2], H. Xu[3], Y. Miao[1], M. Qian[1], S. Wang[1]; [1]Columbia Univ., New York, NY, [2]Brown Univ., Providence, RI, [3]Univ. of Sci. and Technology of China, Hefei, China
Disclosure Block: **T. Wang:** None.
Research in microbiome studies have been centered around the goals to examine whether the entire microbiome community or some taxa are associated with or predictive of a health outcome. Majority of the statistical methods developed for microbiome data are distance-based that begin by computing pairwise distances between microbiome communities of two samples. In these methods, multiple microbiome distance matrices that capture various information in microbiome were examined one by one and the 'optimal' one was selected. That is, only one form of association is used in the final model. However, microbiome studies have suggested that a health outcome usually is associated with multiple taxa in multiple forms. That is, the outcome is associated with some taxa that are phylogenetically-related, while some other associated taxa might be phylogenetically-unrelated, when no single prediction model can capture this complex and mixture forms of association between taxa and an outcome. Here we propose an ensemble prediction method that creates a weighted combination of predictions from three categories of commonly-used prediction methods including i) non-parametric machine learning methods, e.g., random forest (RF), ii) semi-parametric kernel regressions, and iii) linear models. Before ensemble, we propose to enhance RF taking into account the special characteristics of microbiome data through re-weighting contributions of samples using multiple microbial distance metrics in order to capture various aspects of associations between microbiome and a health outcome. We similarly utilize multiple microbial distances in semi-parametric kernel regressions, and then construct the ensemble estimate based on all learners from the three categories. Our simulation studies suggested that the proposed ensemble estimate always has the best prediction performance and a much improved prediction performance than that of any of the ensembled learners when different associations between microbiome and health outcomes were considered. We applied the proposed method to the American Gut Project (EBI: ERP012803) and found that it consistently outperforms competing methods in predicting both binary and continuous microbiome-related health outcomes.

PrgmNr 2689 - An unsupervised approach to identifying haplogroups in mitochondrial DNAs

View session detail

Author Block: **D. Kristjansson**[1,2], T. G. Schurr[3], A. Jugessur[1], J. Bohlin[1]; [1]Norwegian Inst. of Publ. Hlth., Oslo, Norway, [2]Univ. of Bergen, Bergen, Norway, [3]Univ Pennsylvania, Philadelphia, PA
Disclosure Block: **D. Kristjansson:** None.
Introduction: Phylogenetic analysis can provide information about the demography and genetic diversity of a population. Identifying haplogroups in phylogenetic reconstructions of human mitochondrial DNA (mtDNA) sequences is often an iterative task. While the addition of an automated algorithmic program can be used to establish a baseline phylogeny, a phylogenetic tree must often be reconstructed while accounting for known back-mutations and non-continuous variants. The aim of this study was to use a Bayesian Analysis of Population Structure (BAPS) algorithm to identify haplogroups by the partitioning of individual mtDNA sequences into cluster and explore the phylogenetic implications of this clustering.
Methods: hierBAPS is a Bayesian algorithm that assigns a set of haploid DNA sequences to clusters at different levels using an expected total number of groupings for the set of sequences as input, nested within levels over which the clustering should be performed, increasing the total number of groupings. After specifying a number of groupings (K), the algorithm clusters the sequences into as many groupings as possible (up to Kmax groupings). The algorithm then attempts to maximize the posterior probability of an allocation over other possible allocations, generating a data frame that indicates the assignment of sequences to specific clusters. We conducted three different grouping and level combinations using 874 whole mitogenome sequences from haplogroup U5, including 2 levels (4 groupings), 3 levels (11 groupings), and 4 levels (24 groupings) To test whether these groupings were consistent with a maximum likelihood phylogeny, we projected the BAPS clusters onto a phylogeny generated with IQ-tree, which also finds the best nucleotide substitution model. We then predicted haplogroups based on polymorphisms consistent with the most recent version of Phylotree (Build 17).
Results: We found that that using hierBAPS to evaluate human mitogenome sequences accurately predicted the haplogroups for these samples. At the four groupings level, the haplogroups were predicted up to 3-digits. The haplogroup predictions were most optimal at the 24-grouping level, and corresponded to exact haplogroups presented in Phylotree. Through this analysis, we also identified a sublineange of U5b3e1 that had not yet been represented in Phylotree.
Conclusion: Using the hierBAPS algorithm, we defined human mtDNA haplogroups at different resolutions which accurately corresponded to phylogenetically determined clusters. This methodology is also able to identify potenitally new phylogenetic branches that could be added to the known library of mtDNA haplogroups.

PrgmNr 2690 - Combining the strengths of inverse-variance weighting and Egger regression in Mendelian randomization using a mixture of regressions model

View session detail

Author Block: **Z. Lin**[1], Y. Deng[1], W. Pan[2]; [1]Univ. of Minnesota, Minneapolis, MN, [2]Univ Minnesota, Minneapolis, MN
Disclosure Block: **Z. Lin:** None.
With the increasing availability of large scale GWAS summary data on various traits, Mendelian randomization (MR) has become commonly used to infer causality between a pair of traits, an exposure and an outcome. It depends on using genetic variants, typically SNPs, as instrumental variables (IVs). The inverse-variance weighted (IVW) method (with a fixed-effect meta-analysis model) is most powerful when all IVs are valid; however, when horizontal pleiotropy is present, it may lead to biased inference. On the other hand, Egger regression is one of the most widely used methods robust to pleiotropy, but it suffers from loss of power. We propose a two-component mixture of regressions to combine and thus take advantage of both IVW and Egger regression; it is both more efficient (i.e. higher powered) and more robust to pleiotropy (i.e. controlling type I error) than either IVW or Egger regression alone by accounting for both valid and invalid IVs respectively. We propose a model averaging approach and a novel data perturbation scheme to account for uncertainties in model/IV selection, leading to more robust statistical inference for finite samples. Through extensive simulations and applications to the GWAS summary data of 48 risk factor-disease pairs and 63 genetically uncorrelated trait pairs, we showcase that our proposed methods could often control type I error better while achieving much higher power than IVW and Egger regression. We expect that our proposed method will be a useful addition to the toolbox of Mendelian randomization for causal inference.

PrgmNr 2691 - Congenica A.I.: A High Scalable Machine Learning Decision Support Framework for Improving the Diagnostic Yield

View session detail

Author Block: **S. Morganella**, Y. Dahman, L. Ponting, J. R. Filby, N. Trump, H. Savage, E. Serra, S. Drury, A. Haworth, N. J. Lench; Congenica, Cambridge, United Kingdom
Disclosure Block: **S. Morganella:** Salary/Employment; Congenica.
Rare genetic conditions impact more than 400 million people globally with each individual affected by at least one of the ~7,000 recognized rare diseases. Despite the technological advance in DNA sequencing, the characterization of the one or two actionable variants amongst the hundreds of thousands of mutations in an individualâ□□s genome remains a very complex task and hence the identification of an accurate diagnosis for the patient can take several years. One of the main challenges for the clinicians is the lack of enough evidence for delivering a clear classification resulting in variants of uncertain significance (VUS). In this scenario, machine learning can play a key role for delivering reliable and efficient prediction models for the characterization of the diagnostic variants. Here, we present an innovative high scalable machine learning framework that can accurately predict the pathogenicity of the variants, reclassify VUS, and identify the diagnostic variants by using an ad-hoc computational framework that combines patient phenotypes with the information contained in the Online Mendelian Inheritance in Man (OMIM) database. Our model is trained using a unique and well curated set of data generated by expert users of the Congenica platform. The training data includes ~35,000 variants observed in more than 10,000 unique patients and spanning more than 50 rare disease phenotypes. We validated our pathogenicity model on ClinVar 3 and 4 stars variants reviewed by expert panels showing a compelling 95.5% accuracy in predicting the correct pathogenicity. In addition, we showed that we were able to correctly reclassify 92.8% of variants that were previously considered VUS by our clinical team. Finally, we retrospectively assessed our ability to detect the diagnostic variants in 73 singletons, showing that we returned the causative variant as top candidate in 63% of cases and in the top 10 in 85% of cases. We have also analysed 126 singletons obtained from two external research institutes showing that in 48% of cases the diagnostic variant was at the top of the list and in 77% of cases in the top 10. The presented results strongly suggest that by using our machine learning-based decision support framework into clinical settings would help clinicians in increasing the diagnostic yield and improve the diagnosis and treatment even for the most complex cases.

PrgmNr 2692 - Cost-effective Low-Coverage Sequencing Imputation from Millions of Haplotypes

View session detail

Author Block: **S. Rubinacci**[1,2], R. Wertenbroek[1,2,3], O. Delaneau[1,2]; [1]Dept. of Computational Biology, Univ. of Lausanne, Lausanne, Switzerland, [2]Swiss Inst. of Bioinformatics, Lausanne, Switzerland, [3]Univ. of Applied Sci. and Arts Western Switzerland, Yverdon-les-bains, Switzerland
Disclosure Block: **S. Rubinacci:** None.
The last decade saw an unprecedented growth of genetic data with the recent establishment of projects such as TOPMed and GnomAD as a major milestone in human genetics. Despite the significant drop in DNA sequencing costs, SNP arrays are still a common practice to genotype large cohorts of individuals. In the last few years, several studies showed that Low-Coverage Whole Genome Sequencing (LC-WGS) offers a cost-effective alternative to SNP arrays while being less susceptible to ascertainment biases and more powerful to capture rare variations. These results suggest that a shift from SNP arrays to LC-WGS would be beneficial for the next generation of disease and population genetics studies.
In this context, we propose GLIMPSE2, a method for LC-WGS imputation, that is highly accurate while being able to leverage information from millions of reference samples and genetic positions in running times comparable to the last generation of imputation methods designed for SNP arrays.
We show the versatility of our tool by using more than 260 individuals from 127 populations as part of the Simonâ□□s diversity project with coverages ranging from 0.1x to 8x. We report imputation performance at both autosomes (chromosomes 1 to 22) and allosomes (chromosomes X and Y), for different types of small genetic variations (SNPs and indels), and extensively validate that our pipeline offers a solid strategy for genotyping most human populations compared to commonly used SNP array models.
In order to demonstrate the ability of GLIMPSE2 to efficiently leverage information in large reference panels, we performed imputation using simulated data, and showed that GLIMPSE2 is able to scale to reference panels containing millions of haplotypes and hundreds of millions of genetic positions. GLIMPSE2 makes imputation from large reference panels actually so efficient that the main bottleneck now becomes reading files from disk.
Overall, the computational performance reached by GLIMPSE2 makes the tool perfectly suited to perform high throughput imputation from modern high-coverage reference panels in an imputation server setting.

PrgmNr 2693 - Deep Learning in microbiome: Hybrid CNN-LSTM model for disease prediction using longitudinal microbiome data

View session detail

Author Block: **D. Sharma**[1], W. Xu[1,2]; [1]Princess Margaret Cancer Ctr., Toronto, ON, Canada, [2]Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **D. Sharma:** None.
Researchers have shown that changes in the microbiome composition over time, either transiently or long term, through infections or environmental factors can be potentially associated with risk of various diseases. In this work, we propose a hybrid deep learning framework for disease prediction from longitudinal microbiome data, which integrates Convolutional Neural Networks (CNN) for feature extraction and Long Short-Term Memory Networks (LSTM) for analyzing temporal dependency in longitudinal microbiome data along with the hostâ□□s environmental factors. Main points of novelty are: (1) a multi-step approach consisting of step 1: pre-processing microbiome data based on correlation provided by the inherent taxonomic hierarchy in Operational Taxonomic Units (OTUs), to satisfy CNNs requirement for spatial similarity in the input, step 2: incorporating a stratified approach to group OTUs into clusters based on their phylum and training ensemble of CNNs within each of the clusters individually to capture OTU features efficiently and step 3: forwarding the extracted features to the LSTM module at each timepoint to learn temporal dependency; (2) handling irregular observation times and missing data in repeated measures through padding-masking operation in LSTM without the explicit requirement of data imputation and (3) incorporating a weighted loss function in the LSTM learning to handle case-control imbalance, thus, mitigating biased network learning when case-control ratio is highly skewed in microbiome studies. We validated the modelâ□□s performance on a set of 100 simulation studies across multiple time points with drop-off nature in subjects along time in the study design. We further implemented the framework into two real longitudinal human microbiome studies: (i) DIABIMMUNE three-country cohort with four food allergy outcomes (785 samples; 534 OTUs), (ii) DiGiulio study with preterm delivery outcome (3767 samples; 1420 OTUs). Extensive comparison of our proposed model to conventional machine learning methods provided encouraging performance with an Area Under the Curve (AUC) of 0.897 (95% Confidence Interval (CI): [0.891 - 0.905]) on simulated studies and AUCs of 0.762 (95% C.I: [0.758 - 0.765]) on the overall food allergy outcome and 0.713 (95% CI: [0.705 - 0.720]) on the preterm birth outcome for the two real longitudinal microbiome studies showing an improvement of 5%, 19% and 8% respectively, in comparison to the next best performing method. In summary, our analysis provides deeper insights on the efficiency of hybrid CNN-LSTM modelling to understand the associations between longitudinal microbial composition changes and disease outcomes.

PrgmNr 2694 - Detecting differential isoform splicing from single-cell long-read RNA sequencing

View session detail

Author Block: **Y. Hu**[1], K. Wang[2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]CHOP, Philadelphia, PA
Disclosure Block: **Y. Hu:** None.
Unlike conventional single-cell short-read RNA-Seq that only assays gene expression at the 3â end of transcripts, the emergence of single-cell long-read RNA-seq technology has made it possible to measure isoform expression variations at cellular level. This breakthrough enables the investigation of a wider range of research problems including analysis of splicing heterogeneity among individual cells. However, compared to bulk RNA-seq, single-cell long-read RNA-seq data are much noisier due to high technical variability and low sequencing depth. We previously developed LIQA (long-read isoform quantification and analysis) for accurate isoform quantification from long-read bulk RNA-Seq data that accounts for different read (or alignment) quality and different levels of 5â degradation via a survival model. Here we propose LIQA-drop for differential splicing analysis in single-cell long-read RNA-seq, which achieves high sensitivity at low coverage by accounting for technical noises. LIQA-drop explicitly models technical noise by accounting for capture efficiency and amplification bias and further accounts for transcriptional burstiness, which are all unique features encountered in single-cell data. A key aspect of LIQA-drop is a Bayesian-based model to account for prior knowledge of splicing regulation from primary sequence information via deep neural network, making it possible to detect differential isoform usage at low sequencing depth. To evaluate the performance of LIQA-drop, we analyze both simulated and real scRNA-seq datasets. In simulation data, we weaken the true splicing difference between conditions by introducing technical noise through a generative model. We show that LIQA-drop has well-controlled type I error rate (3.2% at 0.05 significance level), and is powerful in detecting differential splicing isoform, especially when splicing difference is small (power = 0.76 when splicing ratio difference = 10%). Also, we show that LIQA-drop accurately estimates isoform usage at cellular level than bulk-level approaches (R^2=0.81). We also apply LIQA-drop to a mouse brain single-cell long RNA-seq dataset with ~1,200 single cells. LIQA-drop identifies unique differential splicing events with subtle difference across 7 cortical cell types, which are consistent with observed junctions reads in IGV sashimi plots. We are currently applying LIQA-drop to multiple real single-cell long-read RNA-seq datasets. With the increasing applications of long-read RNA-seq, we believe LIQA-drop will be well-suited for various isoform splicing studies at cellular level.

PrgmNr 2695 - Distinguishing different psychiatric disorders using polygenic prediction

View session detail

Author Block: **W. Peyrot**[1], A. L. Price[2]; [1]Amsterdam UMC, location VUmc, Amsterdam, Netherlands, [2]Harvard Sch Pub Hlth, Boston, MA
Disclosure Block: **W. Peyrot:** None.
Despite great progress on methods for case-control polygenic prediction (e.g. schizophrenia vs. control), there remains an unmet need for a polygenic method that distinguishes different psychiatric disorders (e.g. schizophrenia (SCZ) vs. bipolar disorder (BIP) vs. depression (MDD) vs. control); such a method would have key clinical value, as symptoms of psychiatric disorders are often shared, in particular at disorder onset. Here, we present a new method for distinguishing different disorders using polygenic prediction.
Our method computes probabilities of each possible diagnosis (e.g. SCZ=60%, BIP=20%, MDD=10%, control=10%) using case-control polygenic risk scores (PRS) for each disorder (computed using existing methods) and prior clinical probabilities for each diagnosis. First, we transform the case-control PRS to their respective liability scales. Second, we use bivariate LD score regression to estimate the genetic correlation among disorders and sample-overlap across the respective discovery GWAS samples. Then, we include all of this information in a multivariate liability threshold model to compute probabilities of each possible diagnosis. To assess prediction accuracy, we compute for each diagnosis the AUC for distinguishing this diagnosis (e.g. SCZ) versus the rest (e.g. BIP, MDD, control) in an independent validation sample.
We simulated data at realistic genetic architectures, genetic correlations and training sample sizes with equal proportions of SCZ, BIP, MDD and controls (25%, 25%, 25% and 25%). In these data, our method yielded well-calibrated results (the predicted probability of each diagnosis was equal to the true probability on average), whereas naÃ¯ve approaches were poorly calibrated. Furthermore, our method attained AUC for distinguishing each diagnosis vs. rest that were comparable to the AUC of published methods for distinguishing cases vs. controls of each disorder, an easier task: 0.70 for SCZ vs. rest, 0.61 for BIP vs. rest, 0.63 for MDD vs. rest, and 0.63 for control vs. rest. These simulation results suggest that our new method could be clinically useful for differential diagnosis, particularly as training sample sizes increase. We will present results of application of the method to empirical psychiatric disease data sets.

PrgmNr 2696 - Exploiting public GWAS databases to identify and adjust for heritable confounding in Mendelian randomization studies

View session detail

Author Block: **J. Morrison**; Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **J. Morrison:** None.
Mendelian randomization (MR) is a widely used causal inference method which promises tests of causal effects derived only from genetic associations. When applied carefully, MR can yield valuable insight. However, pleiotropy and heritable confounding remain common sources of MR false positives. The most reliable MR studies rely on pre-existing evidence and expertise to identify and control for possible sources of confounding. However, for many promising applications of MR, little prior information is available. In lieu of such information, we propose an automated method that exploits the deep and growing resources of publicly available GWAS summary statistics. Our method identifies potential confounders by querying the IEU Open GWAS database for all other known phenotype associations with candidate MR instruments. We then identify traits that explain substantial heterogeneity in per-instrument MR estimates. Our method returns to the investigator a list of potential confounders as well as adjusted estimates of the causal effect. The investigator can then apply their domain knowledge to accept or reject potential confounders, or propose new causal hypotheses. This strategy can serve a dual role of providing more robust MR results and suggesting new candidates for further investigation. We demonstrate an application of this method to an exploration of the relationship between obesity and heart disease, hypertension, and type 2 diabetes.

PrgmNr 2697 - Gene selection by incorporating genetic networks into case-control association studies

View session detail

Author Block: **X. Cao**[1], X. Liang[2], S. Zhang[1], Q. Sha[1]; [1]Michigan Technological Univ., Houghton, MI, [2]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN
Disclosure Block: **X. Cao:** None.
There is strong evidence shown that when genes are functionally related to each other in a genetic network, statistical methods utilizing prior biological network knowledge can outperform other methods that ignore the genetic network structures. Therefore, statistical methods that can incorporate genetic network information into association analysis in human genetic association studies have been widely used since 2008. The most recent published work in this area is by Kim K and Sun H [BMC Bioinformatics]. They developed an approach to incorporate genetic networks into case-control association studies with high-dimensional DNA methylation data. Given a biological network, their approach combines one of data dimension reduction techniques with network-based regularization to identify outcome-related genes. Motivated by their work, we propose a gene selection approach by incorporating genetic networks into case-control association studies with DNA sequence data or DNA methylation data. Instead of using traditional dimension reduction techniques such as principal component and supervised principal component, we use a linear combination of genotypes at SNPs or a linear combination of methylation values at CpG sites in each gene to capture gene-level signals. We develop three approaches for the linear combination: optimally weighted sum (OWS), LD-adjusted polygenic risk score (LD-PRS), and beta-based weighted sum (BWS). OWS and LD-PRS are supervised approaches that depend on the effect of each SNP or CpG site on the case-control status, while BWS can be extracted without using the case-control status. After we use one of the linear combinations of genotypes or methylation values in each gene to capture gene-level signals, we regularize them to perform gene selection based on the biological network. Simulation studies show that the proposed approaches have higher true positive rates than using traditional dimension reduction techniques. We also apply our approaches to Illumina HumanMethylation450 BeadChip array data for DNA methylation and UK Biobank data for DNA sequence of rheumatoid arthritis patients and normal controls. The results show that the proposed methods can select potentially rheumatoid arthritis related genes that are missed by existing methods.

PrgmNr 2698 - Genetic fine-mapping with dense linkage disequilibrium blocks

View session detail

Author Block: **C. Mo**; Univ. of Maryland Sch. of Med., Columbia, MD
Disclosure Block: **C. Mo:** None.
Background: Fine-mapping is an analytical step for causal prioritization of the polymorphic variants in a trait-associated genomic region observed in genome-wide association studies (GWAS). Prioritization of causal variants can be challenging due to linkage disequilibrium (LD) patterns among hundreds to thousands of polymorphisms associated with a trait. Hence, we propose an L0 graph norm shrinkage algorithm to disentangle LD patterns by dense LD blocks consisting of highly correlated single nucleotide polymorphisms (SNPs). We further incorporate the dense LD structure for fine-mapping. From graph theory, the concept of 'dense' refers to that a block is composed mainly by SNPs highly correlated with each other. We demonstrated the application of our new fine-mapping method using a large UK Biobank (UKBB) sample related to nicotine addiction. We also evaluated and compared its performance with existing fine-mapping algorithms using simulations. Results: Our results suggested that polymorphic variances in both neighboring and distant variants can be consolidated into dense blocks of highly correlated loci. Results from simulations demonstrated that this method outperformed comparable fine-mapping methods with increased sensitivity and reduced false-positive error rate for causal variant selection. The application of this method to the smoking severity trait in the UKBB sample replicated previously reported loci and suggested the causal prioritization of genetic effects on nicotine dependency.Conclusion: We found that the dense LD block structure can guide fine-mapping and accurately determine a parsimonious set of potential causal variants. Our approach is computationally efficient and allows fine-mapping of thousands of polymorphisms.

PrgmNr 2699 - Identifying clinically-relevant circulating protein biomarkers for type 1 diabetes: A two-sample Mendelian randomization study

View session detail

Author Block: **N. YAZDAN PANAH**[1], M. Yazdanpanah[1], Y. Wang[2], V. Forgetta[2], M. Pollak[2,3,4], C. Polychronakos[5,6,7], J. Richards[2,3,6,8,9], D. Manousaki[1,10]; [1]Res. Ctr. of the Sainte-Justine Univ. Hosp., Univ. of Montreal, Montreal, QC, Canada, [2]Lady Davis Inst. for Med. Res., Jewish Gen. Hosp., Montreal, QC, Canada, [3]Dept.s of Med., McGill Univ., Montreal, QC, Canada, [4]Dept.s of Oncology, McGill Univ., Montreal, QC, Canada, [5]Dept. of Pediatrics, McGill Univ., Montreal, QC, Canada, [6]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, [7]Ctr. of Excellence in Translational Immunology (CETI), Montreal, QC, Canada, [8]Dept. of Epidemiology and Biostatistics, McGill Univ., Montreal, QC, Canada, [9]Dept. of Twin Res., King's Coll. London, London, United Kingdom, [10]Dept.s of Pediatrics, Biochemistry and Molecular Med., Univ. of Montreal, Montreal, QC, Canada

Disclosure Block: **N. Yazdan panah:** None.

Type 1 diabetes is a chronic autoimmune disorder causing destruction of the insulin-producing pancreatic beta cell. Other than HLA haplotypes and polygenic risk scores, very few biomarkers exist that could predict development of type 1 diabetes before pancreatic auto-antibodies emerge. Such biomarkers might inform us on the pathophysiology of loss of self-tolerance to type 1 diabetes and serve as tools for early screening for individuals at high-risk for the disease. A relatively underexplored source of biomarkers is circulating proteins in the blood, which are drug-binding sites and as such may also represent potential drug targets. Here, we aimed to identify circulating proteins causally linked to type 1 diabetes susceptibility using Mendelian randomization (MR). We employed a large-scale two-sample MR study to screen circulating proteins outside the major histocompatibility complex (MHC) for causal association with risk of type 1 diabetes. To do this, we used cis genetic determinants of up to n=1,611 circulating proteins from five large genome-wide association studies to estimate their causal effects on type 1 diabetes risk amongst 9,684 cases with type 1 diabetes and 15,743 controls. We found that a standard deviation increase in genetically predicted Signal Regulatory Protein Gamma (SIRPG) level was associated with increased risk of type 1 diabetes risk (MR OR = 1.66, 95% 1.36- 2.03; P = 7.1 x 10-7). The risk of type 1 diabetes increased almost two-fold per standard deviation increase in genetically predicted interleukin-27 Epstein-Barr Virus Induced 3 (IL27-EBI3) protein levels (MR OR=1.97, 95% CI = 1.48 - 2.62, P= 3.7 x10-6). Conversely, a standard deviation increase in genetically predicted chymotrypsinogen B1 (CTRB1) was associated with decreased risk of type 1 diabetes (MR OR=0.84, 95% CI = 0.77 - 0.90, P= 6.1 x10-6). In summary, we identified three novel protein biomarkers associated with type 1 diabetes risk using an unbiased MR approach. These biomarkers inform us on the pathophysiology of type 1 diabetes, and, if validated in cohorts of individuals with type 1 diabetes and controls, they may represent promising targets for development of drugs and enhance screening for individuals at high risk for type 1 diabetes.

PrgmNr 2700 - Incorporating family disease history and controlling case-control imbalance for population based genetic association studies

View session detail

Author Block: **Y. Zhuang**[1], B. N. Wolford[2], K. Nam[3], W. Bi[1,4], W. Zhou[5], C. J. Willer[2], B. Mukherjee[1], S. Lee[1,3]; [1]Biostatistics, Univ. of Michigan, Ann Arbor, MI, [2]Univ. of Michigan, Ann Arbor, MI, [3]Seoul Natl. Univ., Seoul, Korea, Republic of, [4]Dept. of Med. Genetics, Sch. of Basic Med. Sci., Peking Univ., Beijing, China, [5]Massachusetts general Hosp., Broad Inst., Boston, MA
Disclosure Block: **Y. Zhuang:** None.
In the genome-wide association analysis of population-based biobanks, the low prevalence in most diseases results in low detection power. If a family disease history is collected, the power can be improved by inferring the disease risks in control samples, which will be used as phenotypes in association analysis. In the presence of related samples, however, the existing methods, such as LT-FH, fail to address increased phenotypic correlation among closely related samples due to similar family history. In addition, existing approaches cannot adjust for the unbalanced phenotypic distribution. We propose a new method, TAPE (mixed-model-based Test with Adjusted Phenotype and Empirical saddlepoint approximation), which controls for increased phenotype correlation by introducing an additional variance component for closely related samples and accounts for case-control imbalance by using empirical saddlepoint approximation. We show through simulation studies that TAPE is computationally efficient and gains greater power than common GWAS without using family disease history (SAIGE) while controlling type I error. In power simulation, TAPE showed 21.0% increase in average chi-square statistics and 12.1% increase in causal SNP detection than SAIGE. While LTFH also had increased power over SAIGE, it suffered type I error inflation especially when analyzing related samples with low disease prevalence and MAF (118 times inflation at alpha=5E-8). We applied TAPE to 10 binary traits in UK Biobank among 408,898 white British samples and identified 659 genome-wide significant clumped variants, among which 127 were with MAF

PrgmNr 2701 - Inter-cohort heterogeneity significantly undermines fine-mapping a meta-analysis

View session detail

Author Block: **M. Kanai**[1,2], R. Elzur[1,2], Global Biobank Meta-analysis Initiative, M. J. Daly[2,1,3], H. K. Finucane[1,2]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Inst. for Molecular Med. Finland (FIMM), Helsinki, Finland
Disclosure Block: **M. Kanai:** None.
Meta-analysis is commonly used to combine multiple genome-wide association studies (GWAS) and has identified thousands of loci associated with complex traits. To resolve causal variants for these loci, previous meta-analysis studies have applied fine-mapping methods as if they are applied to a single-cohort study. However, it is understudied how heterogeneous characteristics of each cohort (e.g., phenotyping, genotyping, or imputation) affect fine-mapping. Indeed, due to such inter-cohort heterogeneity, we observed spurious fine-mapped variants in the recent COVID-19 Host Genetics Initiative meta-analysis when we naively applied existing methods.

Here, we first demonstrate the effect of inter-cohort heterogeneity in meta-analysis fine-mapping via simulations. Using HAPGEN2 with the 1000 Genomes Project Phase3 (1000GP3), we simulated multiple cohorts (n = 10,000 each) from different 1) ancestries (African, East Asian, European, or South Asian); 2) genotyping arrays (Illumina Omni 2.5, MEGA, GSA, or Affymetrix UKB); and 3) imputation panels (1000GP3_hg19, HRC_hg19, or TopMed_hg38). We assigned true phenotypes by assuming the same causal variants and effect sizes across cohorts where each locus has a single causal variant. We then conducted GWAS, meta-analyzed five random cohorts, applied the Approximate Bayes Factor, and evaluated fine-mapping accuracy.

We observed fine-mapping calibration was significantly undermined when we include different genotyping arrays or imputation panels. When multiple genome builds exist, we observed a particularly high number of false positives due to liftover (mean false positive rate: ~18%). With the same array and imputation panel, ancestry does not affect calibration. However, including Africans significantly improves power compared to other ancestries, likely due to shorter length of linkage disequilibrium (LD).

To identify likely problematic loci for meta-analysis fine-mapping, we develop a statistical method using summary statistics and gnomAD LD matrices. Briefly, because marginal association statistics for tagging variants correlate with r2 values to a causal variant, our method regresses chi-squared statistics on r2 to the index variant and identifies outlier variants with unexpected chi-squared, potentially due to inter-cohort heterogeneity. We apply it to 14 disease endpoints from the Global Biobank Meta-analysis Initiative and use the results to guide its fine-mapping effort. Although a principled solution requires complete synchronization across cohorts, our study provides the best practices in meta-analysis fine-mapping and will help prevent spurious results in the future.

PrgmNr 2702 - Joint analysis model of multiple high-throughput sequencing data

View session detail

Author Block: **Y-S. Kim**[1], G. D. Johnson[2], T. Cowart[1], A. S. Allen[2], A. Ochoa[1], A. Barrera[1], T. E. Reddy[1]; [1]Duke Univ., Durham, NC, [2]Duke Univ, Durham, NC
Disclosure Block: **Y. Kim:** None.
Functional genomic assays are commonly used to identify shared and distinct gene regulatory states between diverse collections of cell types and cell states. However, identifying shared regulatory states across the genome becomes challenging as the number of different cell types or cell states becomes large, for example the number of tissues assayed by RNA-seq in the Genotype-Tissue Expression (GTEx) project or the number of different cell models assayed in ENCODE. In particular, as the number of samples increases, the traditional condition-by-condition analysis do not reliably capture commonalities across many samples. To address the challenge of identifying shared gene regulatory states or responses, we developed JAMMYâ a Joint analysis model of multiple biology data. JAMMY is a statistical model that jointly analyzes functional genomics assays across many different cell types and cell states, and identifies responses common and unique to those cells. We demonstrate that JAMMY is highly flexible, able to identify common and unique responses across collections of RNA-seq, ChIP-seq, and even high throughput reporter assay datasets such as STARR-seq.Specifically, we demonstrate that JAMMY can identify cell-shared gene regulatory responses among groups of tissues in GTEx RNA-seq and novel co-binding between transcription factors in ENCODE ChIP-seq data. We also observe, by applying JAMMY to STARR-seq data, our model can identify differences in DNA sequences that contribute to cell-shared or cell-specific regulatory states, and gene regulatory elements that had shared and unique responses to diverse steroid hormones and that contributed to T cell differentiation. We also demonstrate that a joint analysis approach increases statistical power over the traditional pairwise comparisons to discover novel gene regulatory events. Together, these outcomes highlight substantial advantages to joint analysis models for integrating genomic datasets across many cell models.

PrgmNr 2703 - Leveraging Machine Learning (ML) to Predict Inherited Variants Associated with Chronic Lymphocytic Leukemia (CLL)

View session detail

Author Block: **R. Mwangi**[1], G. Kleinstern[1,2], D. Robinson[1], N. J. Boddicker[1], K. G. Rabe[1], A. D. Norman[1], H. Yan[1], C. M. Vachon[1], T. G. Call[3], C. Bonolo de Campos[4], W. Ding[3], J. R. Cerhan[1], N. E. Kay[3], E. Braggio[4], S. L. Slager[1,3]; [1]Dept. of Quantitative Hlth.Sci., Mayo Clinic, Rochester, MN, [2]Sch. of Publ. Hlth., Univ. of Haifa, Haifa, Israel, [3]Div. of Hematology, Mayo Clinic, Rochester, MN, [4]Dept. of Hematology and Oncology, Mayo Clinic, Phoenix, AZ
Disclosure Block: **R. Mwangi:** None.
Background: The current analytical paradigm for genome-wide association studies (GWAS) is to evaluate each single nucleotide polymorphism (SNP) with a specific disease risk. Although this method has successfully identified susceptibility SNPs across cancers, it does not consider possible multi-way and non-linear SNP-SNP interactions. ML algorithms can incorporate interactions without inflating the multiple testing issue by adopting advanced predictive modelling approaches. Herein, we propose to apply these techniques to our existing Mayo GWAS data of patients with CLL to compare ML risk prediction models.
Objective: This project aims to implement advanced ML approaches to GWAS of CLL to compare ML prediction models.
Methods: We genotyped 765 CLL cases and 639 controls with GWAS data from the Mayo Clinic CLL study. Using these data, we implemented a two-step approach by first applying feature selection, which narrows down to the SNPs that have the potential to be highly predictive for CLL risk, followed by a classification step that trains the ML algorithms to classify whether the selected SNPs are associated with CLL risk. We evaluated three types of ensemble algorithms: random forest (RF), extreme gradient boosting machines (XGBoost), and light gradient boosting machines (LightGBM). The performance of their prediction probability was evaluated using the area under the curve (AUC). The more robust algorithm in selecting relevant risk-associated SNPs was chosen. In the second step, we implemented ML classifiers to handle high dimensional data in performing predictions. We evaluated two classifiers, support vector machines (SVM) and regularized logistic regression (RLR) algorithms, on the selected SNPs.
Results: Our feature selection approach showed that LightGBM outperformed both RF and XGBoost and the feature importance score indicated that LightGBM was robust in selecting SNPs with high-risk predictive potential for developing CLL. Algorithm comparisons showed that integrating LightGBM with SVM gives a higher AUC of 63.4% with radial basis function kernel versus an AUC of 60.4% at the baseline SVM, while integrating LightGBM with RLR obtains a higher AUC of 63.6% with elastic-net penalty term when compared to the baseline RLR AUC of 59.4%.
Conclusion: Our preliminary results showed promise that integrating ML techniques for feature selection and classification improves the predictive performance when compared with baseline ML techniques. Our next steps are to incorporate another set of CLL GWAS data of 1,100 CLL cases and 15,000 controls from the Mayo Clinic CLL study and compare the ML findings to that obtained known CLL loci.

PrgmNr 2704 - Leveraging summary statistics fromEuropean GWAS to improve discovery of non European populations accounting for heterogeneity in effect sizes

View session detail

Author Block: **P. Orozco del Pino**[1], S. Zollner[2]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ Michigan, Ann Arbor, MI
Disclosure Block: **P. Orozco del Pino:** None.
Polygenic risk scores (PRS) aim to quantify disease risk that non-genetic tools can not identify raising interest in clinical application These PRS are based on a weighted sum of the number of risk variants found on their genome; weights and risk variants are identified using summary statistics from Genome-Wide Association Studies (GWAS). Since most GWAS samples are of European ancestry, current PRS have lower predictive ability for populations more genetically distant from the European, which means that health disparities will increase if diagnosing and treating tools are based on current PRS.One of the factors for the lack of transferability of PRS is the heterogeneity of effect sizes that arises from differential environmental exposure distributions. However, knowledge of the right environment variables is rare and challenging. In addition, heterogeneity might not be uniform across the genome, which poses the need to account for heterogeneity per region. Finally, combining summary statistics from different populations without accounting for heterogeneity can be more detrimental than beneficial.We develop a method to leverage information from the European sample to reduce the MSE in the estimation of effect size in the non-European sample. To manage the risk of using different populations as discovery and target populations is that the large sample size of the discovery population might hinder the estimation of population-specific effect sizes, we use a power prior distribution to calibrate the discovery population's influence on the posterior distribution of the effect size in the target population. We parametrize this prior based on LD patterns and heterogeneity. We present a model to combine European summary statistics and non-European individual-level data for continuous traits that uses a Bayesian sparse linear mixed model and incorporates the power prior. By modeling the heterogeneity at a variant granularity, we claim that we can increase the estimation of shared effect sizes without compromising the discovery of population-specific effect sizes, which will improve the predictive ability of PRS.

PrgmNr 2705 - M-DATA: A statistical approach to jointly analyzing *de novo* mutations

View session detail

Author Block: **Y. Xie**[1], M. Li[1], W. Dong[2], W. Jiang[1], H. Zhao[1]; [1]Dept. of Biostatistics, Yale Univ., New Haven, CT, [2]Dept. of Genetics, Yale Univ., New Haven, CT
Disclosure Block: **Y. Xie:** None.
Recent studies have demonstrated that multiple early-onset diseases have shared risk genes, based on findings from *de novo* mutations (DMNs). Therefore, we may leverage information from one trait to improve statistical power to identify genes for another trait. However, there are few methods that can jointly analyze DNMs from multiple traits. In this study, we develop a framework called M-DATA (Multi-trait framework for *De novo* mutation Association Test with Annotations) to increase the statistical power of association analysis by integrating data from multiple correlated traits and their functional annotations. Using the number of DNMs from multiple diseases, we develop a method based on an Expectation-Maximization algorithm to both infer the degree of association between two diseases as well as to estimate the gene association probability for each disease. We apply our method to a case study of jointly analyzing data from congenital heart disease (CHD) and autism. Our multi-trait model was able to identify 23 genes including 12 novel genes, which is substantially more than single-trait analysis. Among the 12 novel genes, our analyses suggest *POGZ*, *KDM5B* and *NAA15* may be considered as new candidate CHD genes, leading to novel insights into CHD disease etiology.

PrgmNr 2706 - Modeling transcription factor binding in statistical tests of common and rare non-coding regulatory variants

View session detail

Author Block: **N. Patel**[1], b. jin[2], W. S. Bush[3]; [1]Inst. for Computational Biology, Case Western Reserve Univ, Cleveland, OH, [2]case western reserve Univ., cleveland, OH, [3]Case Western Reserve Univ, Cleveland, OH

Disclosure Block: **N. Patel:** None.

Non-coding variants altering transcription factor(TF) binding can significantly affect gene expression regulation, which has been shown to be associated with complex diseases. Despite the development of transcriptome-wide association study(TWAS) approaches, integrating TF binding information into models of variant influence on gene expression is limited and the inclusion of rare variants (MAF cis-regulatory TF scores for common variants to generate TWAS gene expression prediction models. TFKin is a kernel-based variance component test for performing association between a phenotype and an estimated *cis*-regulatory genetic kinship matrix generated using rare non-coding variant-TF scores. To validate these approaches, we used whole blood genotype and expression data from the Depression Genes Network (N = 922) and GTEx(N =313). We compared the performance of TFXcan with EpiXcan, a TWAS approach that uses broad epigenetic priors generated using eQTL summary statistics to weight variants. We obtained a statistically significant percent improvement(PI) in average cross-validation R^2 for the expression, across all the genes, for TFXcan models compared to the EpiXcan models(PI_{DGN} = 5%, PI_{GTEx} = 10%). This improvement was also seen in average prediction R^2 , when we used models trained using one dataset to impute expression for the other($PI_{DGN â¶ GTEx}$ = 19%; $PI_{GTEx â¶ DGN}$ = 8%). The TFKin approach, besides producing a power of 87%, generated higher percent enrichment(PE) of significant genes replicating in an independent dataset ($PE_{DGNâ¶GTEx}$ = 45%; $PE_{GTExâ¶DGN}$ = 83%), compared to a SKAT based framework weighting rare variants based on their MAF ($PE_{DGNâ¶GTEx}$ = 41%; $PE_{GTExâ¶DGN}$ = 80%). Both, TFXcan and TFKin take advantage of biologically relevant TF based regulatory information to estimate genetic variant influence over gene expression. Furthermore, these methodologies can be easily applied to find genes dysregulated in any complex disease trait.

PrgmNr 2707 - Optimising Polygenic Score Portability Across Ancestral Populations

View session detail

Author Block: **O. Pain**[1], T. Kono[1], R. Arathimos[1], J. R. Coleman[1], K. B. Hanscombe[1], D. Howard[1], G. D. Breen[1], L. Folkersen[2], C. M. Lewis[1]; [1]King's Coll. London, London, United Kingdom, [2]Danish Natl. Genome Ctr., Copenhagen, Denmark
Disclosure Block: **O. Pain:** None.
Introduction: The predictive utility of polygenic scores (PGS) is increasing as the sample sizes within genome-wide association studies (GWAS) become larger and as novel PGS methodology is developed. However, the vast majority of GWAS are based on European ancestry individuals, and PGS have a low portability when applied to non-European populations (portability = $R^2_{\text{non-European}} / R^2_{\text{European}}$). In this research, we compared the portability of PGS across polygenic score methods, and assessed how similarity in allele frequency and linkage disequilibrium (LD) patterns might improve portability. **Methods**: We compare PGS methodology using Height and Body Mass Index (BMI) in unrelated UK Biobank participants (N = 394,434) genetically mapped to the five 1000 Genomes super populations. PGS were calculated using European-based GWAS summary statistics for Height (Wood *et al*. 2014) and BMI (Lock *et al*. 2015). We used previously developed PGS methods, including pT+clump, LDpred2, lassosum, PRScs, SBayesR, and DBSLMM, as well as two novel approaches called imputed gene-expression risk scoring (GeRS), and cross-population similarity weighted polygenic scoring (CrossPop). GeRS integrates multi-SNP predictors of gene expression in the PGS. CrossPop PGS reweight genetic effects according to their similarity in allele frequency and linkage disequilibrium (LD) across populations. **Results**: Previously developed PGS methods showed a similar portability across populations, with relative performance of each method being similar in European and non-European populations. The best prediction was achieved using models containing LDPred2 or lassosum PGS based on a range of shrinkage parameters. Both novel methods, GeRS and CrossPop PGS, provided substantially improved prediction in specific scenarios. For example, using GeRS for BMI in East Asians, the variance explained increased from 3.9% to 7.1% ($p=1.02x10^{-7}$), with a corresponding the portability increase from 0.53 to 0.93 compared to Europeans (baseline value 1.0). Using CrossPop, the variance explained increased from 4.3% to 10.1% ($p=6x10^{-9}$) for Height in Admixed Americans, corresponding to a portability increase from 0.42 to 1.05 compared to Europeans. **Conclusion**: These results indicate that commonly applied polygenic scoring approaches have similar values of portability across ancestral populations. Furthermore, these findings suggest integration of functional annotations, such as effects of gene expression, and allele frequency and LD similarity information, can improve PGS portability across populations. We are currently expanding this study to evaluate these PGS methods in a wider context.

PrgmNr 2708 - Ordinal regression with gene-based aggregation tests in the UK Biobank

View session detail

Author Block: **A. Flynn-Carroll**[1], L. D. Ward[2], A. Deaton[1], M. M. Parker[2], R. Hoffing[1], C. Willis[3], G. Hinkle[1], P. Nioi[2]; [1]Alnylam Pharmaceuticals, Cambridge, MA, [2]Alnylam, Cambridge, MA, [3]Alnylam Pharmaceuticals, Boston, MA
Disclosure Block: **A. Flynn-Carroll:** Salary/Employment; Alnylam Pharmaceuticals.
Gene-based variant aggregation is an important strategy to increase power to detect associations between rare variants and phenotypes of interest. Traditionally these methods have been used when testing for associations with either categorical or quantitative traits. Here we perform burden testing with regression on ordinal traits. Much of the data in the UK Biobank is not accurately captured by binary or quantitative measurements, including many answers to the touch-screen questionnaire. Performing ordinal regression allows for the inclusion of a wider set of data into burden testing. To demonstrate this method, we tested exome-wide gene-based variant aggregations in 363973 individuals against eight ordinal phenotypes, with a minimum of 100 carriers per phenotype, and found type 1 error to be well controlled. The following phenotypes were tested: chronotype, frequency of insomnia, frequency of daytime napping, alcohol intake frequency, overall self-assessed health rating, relative age of first facial hair, usual walking pace, and frequency of recent lack of interest or pleasure in doing things. We found known associations between *ADH1C* predicted loss of function variants (pLOF) with alcohol intake frequency ($p = 4.1 \times 10^{-24}$, beta = 0.23 SD decrease in drinking) and *PER2* pLOF with chronotype ($p = 2.1 \times 10^{-13}$ beta = 1.17 SD increase in morningness). We also found novel associations between *GIGYF1* pLOF and alcohol intake frequency (p-value = 3.5×10^{-14}, beta= 1.14 SD decrease in drinking), *ANKRD12* with alcohol intake frequency (p-value = 5.8×10^{-11}, beta = 0.70 SD decrease in drinking), usual walking pace (p-value = 4.67×10^{-8}, beta = 0.69 SD decrease in walking pace) and overall health rating (p-value = 1.39×10^{-9}, beta = 0.73 SD decrease in overall health), as well as *KDM5B* with daytime napping frequency (p-value = 1.80×10^{-7}, beta = 0.74 SD increase in daytime napping). All three of these genes were previously reported to associate with cognitive function (Chen et al., presentation at ASHG 2020), and we recently reported *GIGYF1* to associate with type 2 diabetes (Deaton et al., medRxiv 2021). These results highlight the importance of including ordinal phenotypes in burden tests of rare variants.

PrgmNr 2709 - PathWAS reveals biologically relevant pathways in complex traits

View session detail

Author Block: **S. M. May-Wilson**[1], E. Macdonald-Dunlop[1], J. F. Wilson[2], N. Pirastu[1]; [1]Univ. of Edinburgh, Edinburgh, United Kingdom, [2]Univ Edinburgh, Edinburgh, United Kingdom
Disclosure Block: **S.M. May-Wilson:** None.
Rationale: With the aim of understanding the biology underlying the genetics of complex traits and multifactorial disease it is increasingly common to incorporate omics, such as in TWAS, to improve the power of discovery and biological relevance of GWAS. However, these studies could be limited by examining the effects of individual genes acting in isolation and not in the context of broader biological networks. Here we present an alternative and novel method for incorporating the information of multiple genes grouped by pathways, in an attempt to predict pathway functionality and to find associations of pathways with complex traits.
Method: We grouped genes from known biological pathways obtained from databases, such as KEGG, and created polygenic risk scores (PRS) for these genes using SNPs from QTL data. These PRS were created using effect sizes weighted by LD structure, using the statistical method LDpred2. The relative contribution of each gene on overall pathway functionality was estimated by fitting a multivariable Mendelian randomisation (MR) analysis, using the QTL SNPs as exposures against one of 184 proteomics â□□end-pointsâ□□ from the SCALLOP consortium, measured in 26000 individuals. The PRS and MR results were then combined to create an overall pathway PRS, validated in an independent sample. The significant pathway scores were then used to search for associations with traits in UK Biobank via PheWAS. We then also conducted additional PheWAS analyses, with the removal of individual genes from the pathway PRS in order to determine if the pathways were the associated factor rather than any individual gene.
Results: Our method successfully predicted the end-point protein level in 12 pathways. These pathways are primarily immune response pathways, such as NOD-like receptor signalling and Toll-like receptor signalling. From these results, PheWAS identified numerous associations between these pathways and traits recorded in UK Biobank, such as lymphocyte and leukocyte count, as well as height, weight and lung-function traits. Removal of individual gene scores resulted in 8 significant pathway-trait associations, several of which are supported by the existing literature.
Conclusion: Pathway scoring offers the prospect of a more powerful and holistic analysis of GWAS results, with the potential to discover relevant causal pathways for complex traits. By incorporating additional QTL and proteomics datasets, along with improvements made in definition of pathways, in future it may be possible to improve the prediction of pathway functionality for any individual, thus allowing potential examination of pathway relevance in many complex traits and diseases.

PrgmNr 2710 - Penalized mediation models for multivariate data

View session detail

Author Block: **D. J. Schaid**[1], O. Dikilitas[1], J. Sinnwell[1], I. J. Kullo[2]; [1]Mayo Clinic, ROCHESTER, MN, [2]Mayo Clinic, Rochester, MN
Disclosure Block: **D.J. Schaid:** None.
Statistical methods to integrate multiple layers of data, from exposures to intermediate traits to outcome variables, are needed to guide interpretation of complex data sets about which variables are likely contributing in a causal pathway from exposure to outcome. Statistical mediation analysis based on structural equation models provide a general modeling framework, yet they can be difficult to apply to high-dimensional data and they are not automated to select the best fitting model. To overcome these limitations, we developed novel algorithms and software to simultaneously evaluate multiple exposure variables, multiple intermediate traits, and multiple outcome variables. Our penalized mediation models are computationally efficient and simulations demonstrate that they produce reliable results for large data sets. Application of our methods to a study of vascular disease demonstrate their utility to identify novel direct effects of single nucleotide polymorphisms (SNPs) on coronary heart disease and peripheral artery disease, while disentangling the effects of SNPs on the intermediate risk factors including lipids, cigarette smoking, systolic blood pressure, and type2 diabetes.

PrgmNr 2711 - Pleiotropy testing identifies genetic variants with pleiotropic evidence in Alzheimerâ□□s disease

View session detail

Author Block: **N. Lorincz-Comi**[1], X. Zhu[2]; [1]PQHS, Case Western Reserve Univ., Cleveland, OH, [2]Case Western Reserve Univ, Cleveland, OH
Disclosure Block: **N. Lorincz-Comi:** None.
Mendelian randomization (MR) analysis can estimate the causal effect of a given exposure on an outcome while identifying pleiotropic variants among instrumental variables via a t-test. The t-test can then be applied to genome-wide variants to search for associations with both the exposure and outcome given the MR causal estimate, and the associations can be interpreted as potential pleiotropy.
In the present study, we studied the causal contributions of brain regions of interests (ROIs) obtained from brain imaging on Alzheimerâ□□s disease (AD) using GWAS summary statistics. Among 101 ROIs, we identified the top 12 ROIs genetically correlated with AD after strict QCs. We estimated causal effects of the top 12 ROIs genetically correlated with AD and performed genome wide pleiotropy analysis. MR results indicate a positive causal effect of hippocampal volume on AD risk (OR=1.02, P=0.015), and a negative effect of the caudate (OR=0.51, P=0.024) on AD risk. From MR models for all 12 brain regions, one SNP (rs35283920, 16:70687195:T:C, 31.3% MAF) was genome-wide significant in pleiotropy testing but had no genome-wide significant associations in the AD GWAS in a 1 Mb window. As such, rs35283920 tagged a novel locus that was undetected in the original AD GWAS. This SNP had associative evidence (effect size=0.01, P=0.048) in respective ROI GWAS (ventral diencephalon). The gene closest to this SNP, *IL34*, was thus not originally identified by the AD GWAS of >425,000 participants, cases defined as having AD or a family history of AD. However, this gene was identified in a smaller GWAS (nâ□□315,000) of AD family history only, this study including additional clinical covariates in their GWAS analyses.
We demonstrated that the MR pleiotropy t-test can be used to identify novel SNP associations undetected by standard GWAS analysis. The identified associations potentially have pleiotropic effect. We additionally aim to cluster the 101 ROIs into a small number of distinct groups based on genetic correlations and treat these groups as exposures in MR analysis with AD, educational attainment, and cognition.

PrgmNr 2713 - scRegMap: A statistical framework for mapping context-specific regulatory variants using single cell RNA-sequencing

View session detail

Author Block: **A. Cuomo**[1,2], T. Heinen[3], D. Vagiaki[4], D. Horta[2], J. C. Marioni[2,1,5], O. Stegle[3,4,1];
[1]Wellcome Sanger Inst., Hinxton, United Kingdom, [2]European Bioinformatics Inst. (EMBL-EBI), Hinxton, United Kingdom, [3]DKFZ, Heidelberg, Germany, [4]EMBL, Heidelberg, Germany, [5]Cambridge Inst. Cancer Res. UK (CRUK), Cambridge, United Kingdom
Disclosure Block: **A. Cuomo:** None.
Single cell RNA sequencing (scRNA-seq) is widely applied to assess cellular heterogeneity in human tissues and cell-based models. Technological advances and exponential reduction in cost have enabled the first population-scale scRNA-seq studies, which have assayed single-cell transcriptomes in hundreds of genetically diverse individuals. However, current workflows to analyse these data remain largely based on principles for analysing conventional bulk expression quantitative trait locus (eQTL) studies, and hence fail to fully exploit complex scRNA-seq readouts. A critical limitation of current methodology is the need to define cell types for eQTL mapping a priori, which limits novel opportunity to chart continuous and unbiased landscapes of regulatory variants. To address this, we here proposed scRegMap, a statistical framework to map regulatory variants across the continuous manifold of cellular environments estimated from scRNA-seq. scRegMap allows to test for and quantify genetic effects on gene expression at the resolution of individual cells, while flexibly sharing statistical strength related cell states. Our framework provides a principled strategy to identify and characterize heterogeneous genetic effects that vary across cell states and cell types. We validate scRegMap using simulated data and apply it in the context of recent single-cell eQTL studies of differentiating iPS cells from our lab [1,2] where we demonstrate its power to map context-specific genetic effects and identify the cellular subpopulations where they are active. By clustering single-cell allelic effects, we identify regulatory modules as eQTLs with shared response patterns across cell states. Moreover, using our approach, we identify novel colocalization signals that connect a subpopulation of dopaminergic neurons to schizophrenia risk.
[1] Cuomo et al., Nature Communications, 11 (1), 1-14 (2020)[2] Jerber et al., Nature Genetics, 53 (3), 304-312 (2021).

PrgmNr 2714 - Simulated data provides insight on optimal control method for confounders

View session detail

Author Block: **L. Breidenbach**[1], L. K. Davis[2]; [1]Vanderbilt, Maple Grove, MN, [2]Vanderbilt Univ Med Ctr., Nashville, TN
Disclosure Block: **L. Breidenbach:** None.
Case/control studies are a common method to study how an exposure affects an outcome. However, population stratification, environmental variables, and other confounders often obscure results. Different control methods can be employed to correct for these problems. One popular approach is regression, where each confounder is accounted for by weighting its impact on the exposure and outcome. Another popular approach is matching, which matches controls and cases on potential confounding variables. There are many subtypes of analyses within these two general approaches, thus creating a multitude of options for controlling confounding. However, there is little guidance for choosing an appropriate control method for a given data set. Here we hypothesize that the choice of a control method has a significant impact on results, and that knowing which one to choose will lead to more accurate, reproducible science. To test this, we simulated data of various sample sizes (N=100 through N=10,000) using Bayesian Networks (BNs). Th BNs allowed us to embed complex relationships between the exposure, the outcome, and confounders, modeled after directed acyclic graphs (DAGs). Next, we varied the effect strength the exposure had on the outcome. We then tested the ability of various regression and matching methods of confounding control to recover the inputted effect strength. We compared the regression-based control methods to matched control methods to determine which approach yielded the most accurate estimate of the true effect size.

PrgmNr 2715 - Simulations to understand potential miscalibration in fine-mapping algorithms

View session detail

Author Block: **R. Elzur**[1,2], R. Cui[3,2], M. Kanai[1,2], J. C. Ulirsch[1,2], Z. Fan[4], H. K. Finucane[5,1,2]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Analytic and Translations Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [3]Broad Inst, Cambridge, MA, [4]Yale Univ., Dept. of Statistics and Data Sci., New Haven, CT, [5]Harvard T. H. Chan Sch. of Publ. Hlth., Cambridge, MA
Disclosure Block: **R. Elzur:** None.
Genetic fine-mapping uses genome-wide association study results to select variants with a high probability of being causal for a trait. State of the art algorithms such as SuSiE and FINEMAP have led to biological insights; however, we have found that these and other algorithms produce posterior inclusion probabilities (PIP) for variants that differ greatly at different sample sizes. Specifically, when fine-mapping a range of well-powered phenotypes in the UK Biobank using sample sizes of 360k versus 100k individuals, we see a high replication failure rate (RFR) of up to 36%, that is, a high proportion of variants reported at N=100k as putatively causal (PIP>0.9) reported as not putatively causal (PIPWe have re-created replication failure in small-scale simulations. First, in simulations with variant effect sizes drawn from a uniform distribution instead of a normal distribution, as specified by the prior, RFR for SuSiE approached 12% due to false positives at the smaller sample size. Second, in loci with 2 causal variants with similar effect size in linkage disequilibrium r = 0.9, SuSiE and Approximate Bayes Factors RFRs exceeded 10%. However, these and other simulations that achieve RFRs approaching those observed in real data do not align with what are considered realistic genetic architectures for complex traits.

To better understand what might be driving miscalibration, we simulated larger, more realistic data. We drew random genotypes from UK Biobank imputed genotype probabilities. We then sampled causal variants and effect sizes from the SuSiE posterior distributions obtained by fine-mapping genome-wide significant loci for Height in 360k individuals from the UK Biobank. We randomly selected additional causal variants in accordance with polygenic models for complex traits, and we simulated uncorrected population stratification. After simulating phenotypes, we applied quality control thresholds and excluded some simulated causal variants. Ongoing work involves conducting an association study, fine-mapping significant loci, and assessing calibration on this data.

PrgmNr 2716 - Spatially Aware Dimension Reduction for Spatial Transcriptomics

View session detail

Author Block: **L. Shang**[1], X. Zhou[2]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ MICHIGAN, Ann Arbor, MI
Disclosure Block: **L. Shang:** None.
Spatial transcriptomics are a collection of genomic technologies that have enabled transcriptomic profiling on tissues with spatial localization information. Analyzing spatial transcriptomic data is computationally challenging, as the data collected from various spatial transcriptomic technologies are often noisy and display substantial spatial correlation across tissue locations. Here, we develop a spatially aware dimension reduction method, SpatialPCA, that can extract a low dimensional representation of the spatial transcriptomics data with enriched biological signal and preserved spatial correlation structure, thus unlocking many existing computational tools previously developed in single-cell RNAseq studies for tailored and novel analysis of spatial transcriptomics. We illustrate the benefits of SpatialPCA for spatial region detection, trajectory inference on the tissue, and high-resolution spatial map construction in multiple spatial transcriptomic datasets. In these applications, SpatialPCA identifies key molecular and immunological signatures in a newly detected tumor surrounding microenvironment including a tertiary lymphoid structure that shapes the gradual transcriptomic transition during tumorigenesis and metastasis, detects the past neuronal developmental history that underlies the current transcriptomic landscape across tissue locations in the cortex and cerebellum, and reveals fine-resolution structural and transcriptomic architectural details underpinning the functionality of complex tissues.

PrgmNr 2717 - Spatially Informed Cell Type Deconvolution for Spatial Transcriptomics

View session detail

Author Block: **Y. Ma**[1], X. Zhou[1,2]; [1]Dept. of Biostatistics, Univ. of Michigan, Ann Arbor, MI, [2]Ctr. for Statistical Genetics, Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **Y. Ma:** None.
Various spatially resolved transcriptomic technologies have enabled gene expression profiling on complex tissues with spatial localization information. The majority of these technologies, however, are currently of limited spatial resolution and effectively measure on each tissue location the average gene expression from a mixture of cells of potentially heterogeneous cell types. Consequently, deconvoluting cell types on tissue locations measured in spatial transcriptomics becomes a key for identifying the spatial localization of different cell types and characterizing the structural and functional organization of complex tissues. Here, we introduce a deconvolution method, CARD, that leverages cell type specific expression information from single cell RNA sequencing (scRNA-seq) for the deconvolution of spatial transcriptomics. A unique feature of CARD is its ability to model the spatial correlation in cell type composition across tissue locations, thus enabling spatially informed cell type deconvolution. Modeling spatial correlation allows us to borrow the cell type composition information across locations on the entire tissue to accurately infer the cell type composition on each individual location, achieve robust deconvolution performance in the presence of mismatched scRNA-seq reference, impute cell type compositions and gene expression levels on unmeasured tissue locations, and facilitate the construction of a refined spatial tissue map with a resolution much higher than that measured in the original study. We demonstrate the benefits of CARD through extensive simulations and in-depth analysis of four spatial transcriptomics data sets with distinct technologies. In the real data applications, CARD revealed refined tissue structures with enhanced spatial resolution and identified novel marker genes that underlie the functional and structural organization of the olfactory bulb and hippocampus. CARD also identified multiple cell types and molecular markers with distinct spatial localization that define the progression, heterogeneity and compartmentalization of pancreatic cancer.

PrgmNr 2718 - Stacking Method to Improve Polygenic Risk Prediction in Admixed Individuals using Population Specific GWAS Effect Sizes

View session detail

Author Block: **K. S. Liao**[1], S. Zollner[2]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Univ Michigan, Ann Arbor, MI

Disclosure Block: **K.S. Liao:** None.

Polygenic risk score (PRS) predictions for admixed individuals, in whom genomic segments trace back to multiple ancestries, perform poorly when constructed from European-ancestry GWAS. Incorporating GWAS effect size estimates from each population represented in a sampleâ□□s admixture may improve polygenic prediction. Here, we propose a machine learning stacking method to improve polygenic risk prediction in admixed individuals that leverages population specific GWAS effect sizes. Our stacking method uses a window-based framework to account for heterogeneity in local admixture across samples. In training data, a local European (EUR) and African (AFR) PRS are computed in each window across the genome via clumping and thresholding (C+T) using their respective population GWAS. We then â□□stackâ□□ local PRSs across windows from the training data in a penalized regression model using all local PRSs as covariates. Stacking in a penalized framework provides an efficient data-driven approach to learn which populationâ□□s effect sizes maximize prediction in a local region. In testing data, learned weights from the penalized regression model are used to compute a stacked PRS as a linear combination of local PRSs. For a baseline comparison, we compute a â□□traditionalâ□□ EUR and AFR PRS using genome wide C+T. To validate our method, we simulate admixed African Americans (AA) and GWAS effect sizes for EURs and AFRs under a shared genetic architecture. Method performance is assessed using the Pearson correlation between the â□□trueâ□□ PRS (PRS using only the designated GWAS causal variants) and estimated PRS for each method. Across simulations, the average overall correlation between the stacked PRSs and true PRSs was 0.79 compared to 0.58 and 0.57 for the traditional EUR and AFR PRSs, respectively. When stratifying by EUR ancestry quantiles, the stacked PRSs outperformed both traditional PRSs with correlation increases to the true PRS ranging from 26.1% to 52.2%. Furthermore, while the performance of the traditional PRSs showed ancestry dependence (correlations for the EUR and AFR PRSs linearly increased by 3.24% and decreased by 1.58% respectively for each EUR ancestry quantile increase), our stacking approach removed the relationship between PRS performance and global ancestry with a mere 0.65% correlation increase across quantiles.

We based our work on simulations to ensure high-powered GWAS effect sizes in the understudied African population. With ongoing efforts to fund larger genetic studies in diverse populations, our approach will allow for a simple data-driven method for polygenic risk prediction in individuals of all admixture types in the near future.

PrgmNr 2719 - Testing Mendelian Randomization methods via zero-relevance analysis of COVID-19 traits

View session detail

Author Block: **T. Bond**[1,2,3,4], T. Palmer[1,4], D. Evans[2,5,6], G. Davey Smith[5,4], D. Lawlor[1,4]; [1]MRC Integrative Epidemiology Unit at the Univ. of Bristol, Bristol, United Kingdom, [2]Univ. of Queensland Diamantina Inst., Brisbane, Australia, [3]Dept. of Epidemiology and Biostatistics, Imperial Coll. London, London, United Kingdom, [4]Population Hlth.Sci., Univ. of Bristol, Bristol, United Kingdom, [5]MRC Intergrative Epidemiology Unit at the Univ. of Bristol, Bristol, United Kingdom, [6]Inst. for Molecular BioSci., The Univ. of Queensland, Brisbane, Australia
Disclosure Block: **T. Bond:** None.
Background: Mendelian randomization (MR) is an increasingly used group of methods which exploit genetic association data to improve causal inference. The validity of MR is undermined when genetic variants affect the outcome independently of the exposure, particularly via confounders of the exposure-outcome relationship (correlated horizontal pleiotropy). The COVID-19 pandemic provides an opportunity to test the robustness of recent MR methods which aim to overcome this limitation. Participants of GWAS conducted prior to 2020 constitute a â□□zero-relevanceâ□□ group, in which COVID-19 traits cannot conceivably have had causal effects on any outcomes. We tested whether MR methods correctly infer this, using COVID-19 traits as negative control exposures. **Methods:** Using Linkage Disequilibrium Score Regression (LDSC) we estimated genetic correlations (rG) between a range of socioeconomic and biomedical traits in the UK Biobank (UKB), and two COVID-19 Host Genetics Initiative phenotypes: hospitalized COVID-19 cases, and cases with COVID-19 or SARS-CoV-2 infection, each versus separate control groups (comprising all non-case individuals). We then applied numerous MR methods in a Two-Sample MR Phenome-Wide Association analysis (MR-PheWAS), testing effects of liability to COVID-19 traits on 869 outcome traits. As a positive control, we conducted an MR-PheWAS in the opposite direction, testing effects of 869 traits on COVID-19 outcomes. **Results:** LDSC provided strong evidence for genetic correlation between COVID-19 traits and numerous socioeconomic-related traits, raising the prospect that COVID-19 traits are subject to pervasive correlated pleiotropy via socioeconomic confounders. MR-PheWAS analyses indicated that the MR-Egger, weighted median and weighted mode-based estimators inferred more statistically significant causal effects of COVID-19 liability on other traits than would be expected by chance. Widely used pleiotropy detection methods did not provide evidence for horizontal pleiotropy. **Conclusions:** Our results suggest that commonly used MR methods may have limited robustness to pleiotropic bias, so should not be viewed as definitive evidence for a causal effect in isolation. COVID-19 traits may provide valuable negative control exposures for studies developing or applying MR methods.

PrgmNr 2720 - The construction of multi-ethnic polygenic risk score using transfer learning

View session detail

Author Block: **Z. Zhao**[1], B. Mukherjee[1], S. Lee[2]; [1]Univ. of Michigan, Ann Arbor, MI, [2]Seoul Natl. Univ., Seoul, Korea, Republic of
Disclosure Block: **Z. Zhao:** None.
Methods for polygenic risk scores (PRS), including pruning and thresholding (PT), lassosum (lsum) and PRS-CS, have been extensively investigated in recent years. As most existing GWAS were conducted in European or East Asian individuals, the existing PRS models have limited transferability to minority populations such as Africans and South Asians. Although recent studies have developed multi-ethnic PRS models that linearly combine multiple PRS trained with different ancestry GWAS, they remain under-powered.
Here we propose a novel multi-ethnic PRS using transfer learning from machine learning literature. Our approach, TL-PRS, fine-tunes the potentially biased model trained with GWAS summary statistics from the majority ancestry to the target dataset of the minority ancestry. TL-PRS can use any existing PRS methods (such as lsum and PRS-CS) as a baseline method for fine-tuninge. Using the potentially biased baseline parameter estimates as initial values, TL-PRS iterates the gradient descent algorithm to adapt the parameters for the target ancestry group. In the presence of multiple GWAS summary statistics from different ancestries, TL-PRS combines fine-tuned PRS using the linear combination. Through simulation studies, we show that TL-PRS improved the performance of PRS with a wide range of genetic architectures and cross-population genetic correlations. TL-PRS was most effective when genetic correlations between populations and samples used for calculating summary statistics discovery GWAS sample sizes were small and the genetic architecture was less polygenic. For example, when genetic correlation was 0.4, TL-PRS-lsum attained on average a 228% and 36% relative improvement in prediction accuracy compared to lsum when the causal variants were 0.1% and 1% respectively, while TL-PRS-CS attained on average a 26% and 20% relative improvement compared to PRS-CS.
In the application of 8,168 Africans of UK Biobank data, TL-PRS substantially improved the prediction accuracy of all six quantitative and two dichotomous traits. Compared to lsum, TL-PRS-lsum attained a 21% and 26% average relative improvement in prediction accuracy when using Biobank Japan and UK Biobank GWAS summary statistics, respectively; the average relative improvement of TL-PRS-CS over PRS-CS was 40% and 9%, respectively. When combining summary statistics from Biobank Japan and UK Biobank, the TL-PRS-lsum and TL-PRS-CS outperformed the linear combination of PT and PRS-CSx. By improving the polygenic risk prediction in non-European individuals, our approach will increase the usefulness of PRS and reduce potential health disparities.

PrgmNr 2721 - The METRO method improves identification of gene-disease associations through multi-ancestry transcriptome-wide association studies

View session detail

Author Block: **Z. Li**[1], W. Zhao[2], L. Shang[1], T. H. Mosley[3], S. L. R. Kardia[2], J. A. Smith[2], X. Zhou[1]; [1]Dept. of Biostatistics, Sch. of Publ. Hlth., Univ. of Michigan, Ann Arbor, MI, [2]Dept. of Epidemiology, Sch. of Publ. Hlth., Univ. of Michigan, Ann Arbor, MI, [3]Memory Impairment and Neurodegenerative Dementia (MIND) Ctr., Univ. of Mississippi Med. Ctr., Jackson, MS
Disclosure Block: **Z. Li:** None.
Transcriptome-wide association studies (TWAS) integrate genome-wide association studies (GWAS) with gene expression mapping studies for identifying gene-trait associations. So far, almost all existing TWAS methods have focused on using expression studies collected on individuals from a single genetic ancestry, typically European ancestry, for building expression prediction models that are necessary for the subsequent TWAS analysis. However, due to the differences in allele frequencies and linkage disequilibrium patterns, SNP effects on gene expression can vary, sometimes quite substantially, across populations with diverse genetic backgrounds. Consequently, the expression prediction models constructed in one genetic ancestry may not necessarily benefit or transfer to another, thus hindering the application of TWAS towards underrepresented populations. In addition, the expression prediction models constructed in different groups of genetic ancestry may capture distinct genetic architectures underlying gene expression and thus may provide complementary information that could otherwise improve the effectiveness of TWAS. Here, we develop a new computational method, METRO (Multi-ancEstry TRanscriptOme-wide analysis), that leverages expression data collected from multiple genetic ancestries to enhance TWAS. METRO incorporates expression prediction models constructed in multiple genetic ancestries through a likelihood-based inference framework, producing calibrated test statistics with substantially improved TWAS power. We illustrate the benefits of METRO through comprehensive simulations and applications to four GWAS datasets that include two of primarily European ancestry and two of primarily African ancestry. In the real data applications, we leverage expression data measured on 1,032 African Americans and 801 European Americans from the Genetic Epidemiology Network of Arteriopathy (GENOA) study to identify a substantially larger number of gene-trait associations as compared to existing TWAS approaches. Among the newly identified associations are *ABCA1* and *SIK3* with high density lipoprotein and *MAPT* with type II diabetes. These new associations highlight the critical role of cellular transportation and homeostasis in regulating lipid metabolism and provide evidence that the pathology in microtubule-associated protein tau may act as a key trigger of impaired insulin sensitivity and secretion in the etiology of type II diabetes.

PrgmNr 2722 - Ultrahigh dimensional learning of polygenic risk scores for Mendelian randomization studies

View session detail

Author Block: **X. Zhang**, D. Kong, L. Wang, S. Volgushev; Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **X. Zhang:** None.

Mendelian randomization (MR) is a statistical method by which genetic variants, usually single nucleotide polymorphisms (SNPs), are leveraged as instrumental variables (IV) to examine the causal relationship between modifiable exposure or risk factor and a clinically relevant outcome from observational data, while control for any unobserved confounder. Finding appropriate genetic variants is crucial to make convincing causal conclusions from MR analysis. Previously, selection of genetic variants replies on expert knowledge. Nowadays, explosion of modern data collection has facilitated access to a larger set of genetic variants associated with risk factors and disease-related outcomes. Current methods work well when candidate instruments are of moderate size. However, for the identification of valid IVs under ultrahigh dimensional setting, normal in practice, empirical evidence implies that existing procedures may miss many or even all the valid IVs, due to inclusion of irrelevant variables that have non-negligible correlation with the exposure.

In response to this challenge, we propose a novel approach to first remove irrelevant variants from the candidate set. Some variants in the remaining set might not be valid instruments due to pleiotropic effects, linkage disequilibrium and so on. To address this problem, we then apply existing methods that aims to make valid causal inference in the presence of invalid IVs to identify valid instruments. To obtain more accurate causal effect estimate, we aggregate the effect of each estimated valid IV by constructing polygenic risk score (PGS), which may explain a considerable proportion of variation in the exposure and thus produce an adequately powered MR analysis. In addition, we provided theoretical guarantee of the proposed procedure.

To evaluate the performance, we investigate the causal relationship between tau protein and Alzheimerâ□□s Disease (AD). The data we use comes from the Alzheimerâ□□s Disease Neuroimaging Initiative (ADNI) database. In our analysis, we consider 357 subjects and approximately 3 million SNPs as candidate instruments. The proposed method has identified a new set of genetic variants that were missed by existing approaches.

PrgmNr 2724 - A gene association study on preterm birth with sensitivity analysis to exposures in a Latin American population

View session detail

Author Block: **D. E. Elias**[1], H. CampaÃ±a[1,2], F. A. Poletta[1,3], S. L. Heisecke[4], J. A. Gili[1,5], J. Ratowiecki[1], M. Pawluk[1], M. R. Santos[1,2,6], V. Cosentino[1,7], R. Uranga[1,8], D. Rojas MÃilaga[9], A. Brinckmann Oliveira Netto[9], A. C. Brusius-Facchin[9], C. Saleme[10], M. Rittler[1,11], H. B. Krupitzki[12], J. S. Lopez Camelo[1,3], L. G. Gimenez[1,3]; [1]Estudio Colaborativo Latino Americano de Malformaciones CongÃ©nitas (ECLAMC), Centro de EducaciÃ³n MÃ©dica e Investigaciones ClÃnicas-Consejo Natl. de Investigaciones CientÃficas y TÃ©cnicas (CEMIC-CONICET), Ciudad AutÃ³noma de Buenos Aires, Argentina, [2]ComisiÃ³n de Investigaciones CientÃficas, Buenos Aires, Argentina, [3]Inst. Natl. de GenÃ©tica MÃ©dica Populacional (INAGEMP), CEMIC-CONICET, Ciudad AutÃ³noma de Buenos Aires, Argentina, [4]DirecciÃ³n de InvestigaciÃ³n, CEMIC-CONICET, Ciudad AutÃ³noma de Buenos Aires, Argentina, [5]Inst. AcadÃ©mico PedagÃ³gico de Ciencias Humanas, Univ. Natl. de Villa MarÃa, CÃ³rdoba, Argentina, [6]Inst. Multidisciplinario de BiologÃa Celular, Buenos Aires, Argentina, [7]Hosp. Interzonal Gen. de Agudos Luisa C. de Gandulfo, Buenos Aires, Argentina, [8]Hosp. San Juan de Dios, Buenos Aires, Argentina, [9]ServiÃ§o de GenÃ©tica MÃ©dica, Hosp. de ClÃnicas de Porto Alegre (HCPA), Porto Alegre, Rio Grande do Sul, Brazil, [10]Inst. de Maternidad y GinecologÃa Nuestra SeÃ±ora de las Mercedes, TucumÃin, Argentina, [11]Hosp. Materno Infantil RamÃ³n SardÃi, Buenos Aires, Argentina, [12]Inst. Univ.rio, Centro de EducaciÃ³n MÃ©dica e Investigaciones ClÃnicas (CEMIC-IUC), Ciudad AutÃ³noma de Buenos Aires, Argentina

Disclosure Block: **D.E. Elias:** None.

Introduction: Preterm birth (PTB) is the main condition related to perinatal morbimortality worldwide. The aim of this exploratory study was to identify associations of genetic variants with spontaneous PTB considering maternal environmental exposures.

Methods: We carried out a retrospective case-control study including sociodemographic, neighborhood, and obstetric data of women who gave birth at a maternity hospital from TucumÃin, Argentina, between 2005 and 2010. Using the Ion Torrent S5 platform, we sequenced the exons of *KCNN3*, *CRHR1*, *COL4A3*, *PON1*, and *F3* genes from blood samples of 69 preterm (case) and 61 term newborns (control). We identified genetic variants and performed quality controls using the Ion Torrent procedure. We studied the association of genetic variants with PTB using two methodologies, in both we used an additive genetic model. On the one hand, we used a penalized logistic regression including variants, ancestry, fetal sex, and 48 variables related to maternal exposures (sociodemographic, neighborhood, diseases, health care, and obstetric). We selected the penalized model based on binomial deviance using leave-one-out cross validation. On the other hand, we performed a multiple correspondence analysis (MCA) on maternal exposure variables. Then we performed logistic mixed models for each variant including covariates related to ancestry, fetal sex, and increasing the inclusion of MCA components from 0 to 10 as covariates. We used Wald test to get effect size estimates and analyzed these models using quantile-quantile plots and genomic control.

Results: In the studied sample we identified 57 genetic variants. The T allele of rs4845397 (*KCNN3*) variant was included in the selected model with the penalized logistic regression while in all logistic mixed models, it had an odd ratio lower than 1 with a 95% confidence interval. The genomic control of the logistic mixed model that included the first 3 MCA components, which explain 19.4% of the variance of maternal exposures, was 1.02. In this model the variant odd ratio was 0.25 (CI 95% 0.10 - 0.62, P 0.0025, false discovery rate 0.14). The models that included 4 or more MCA components reported a genomic control greater than 1.10. In the studied sample the frequency of the T allele was 16.5%.

Conclusions: The results of this exploratory study suggest that the *KCNN3* variant would be

associated with spontaneous PTB considering 19.4% of the variance of maternal exposures. Future studies with a larger sample size are necessary to confirm these findings and to analyze a greater number of exposures.

PrgmNr 2725 - A phenome-wide association study of telomere length using combined qPCR and whole genome sequencing data in 136,597 UK Biobank participants

View session detail

Author Block: **O. Burren**, S. V. V. Deevi, Q. Wang, H. Olsson, K. J. Carss, A. S. Platt, S. Petrovski; AstraZeneca, Cambridge, United Kingdom
Disclosure Block: **O. Burren:** Major Stockholder/Ownership Interest; astrazeneca. Salary/Employment; astrazeneca.
Variation in telomere length (TL) between individuals has been associated with a range of human diseases, including cancer and idiopathic pulmonary fibrosis (IPF). Previous population-based studies have successfully used either qPCR[1] or whole-genome sequencing[2] (WGS) measurements to characterise phenotypic and genetic associations with TL. In this pilot study we compared qPCR and WGS TL estimates for 136,597 UK Biobank (UKB) participants, finding only partial correlation (r=0.41) between the two measures. We hypothesised that a PCA-derived linear combination[3] of both WGS and qPCR measurements might optimise biological signal and boost power for phenotypic and genetic association analyses. For the same ~137K UK Biobank participants, we compared GWAS statistics of our combined PCA TL score with those derived from qPCR and WGS measurements, identifying 64, 55 and 42 common variant loci, respectively. And, in support of the combined PCA TL approach all but two of the 64 loci from our PCA TL score replicated in a larger ~470,000 qPCR-only UKB study[1]. Using our PCA TL score we examined over 33,000 binary clinical endpoints finding 106 significant associations that included known positive (e.g., prostate cancer $P_{pc1}=1.7 \times 10^{-15}$, $\hat{I}^2_{pc1}=1.6$) and negative (e.g., IPF $P_{pc1}=7.8 \times 10^{-43}$, $\hat{I}^2_{pc1}=-0.6$) relationships with TL. By applying a rare variant collapsing approach, we identified ten genes that when genetically aberrated significantly associated with decreased or increased TL. These included three genes where carrying a rare protein-truncating variant associated with maintaining a significantly increased TL compared to the non-carriers. Our pilot study highlights not only the utility of WGS to robustly measure TL at scale but also the value of combining orthogonal TL measurements to demonstrably increase power in downstream analyses. Ongoing work will leverage the WGS and qPCR TL measurements of over 450,000 UKB participants, serving as a platform to both define novel telomere biology genetics and connect this with underlying disease pathogenesis.

1. Codd,V. *et al.* (2021). *MedArxiv http://dx.doi.org/10.1101/2021.03.23.21253516*
2. Taub,M.A. *et al.* (2020). *BioArxiv http://dx.doi.org/10.1101/749010*
3. Aschard,H. et al. (2014). Am. J. Hum. Genet., 94, 662-676.

PrgmNr 2726 - A tissue-level phenome-wide network map of colocalized genes and phenotypes in the UK Biobank

View session detail

Author Block: **G. Rocheleau**[1,2], I. S. Forrest[1,2,3], A. Duffy[1,2], S. Bafna[1], A. Dobbyn[1,2,4], M. Verbanck[5], H-H. Won[6], D. M. Jordan[1,2], R. Do[1,2]; [1]The Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Med. Scientist Training Program, Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Pamela Sklar Div. of Psychiatric Genomics, Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]FacultÃ© de Pharmacie de Paris, UniversitÃ© de Paris, Paris, France, [6]Sungkyunkwan Univ., Seoul, Korea, Republic of

Disclosure Block: **G. Rocheleau:** None.

Background Phenome-wide association studies conducted in electronic health record (EHR)-linked biobanks have uncovered a large number of genomic loci associated with traits and diseases. However, interpretation of the complex relationships of associated genes and phenotypes is challenging. **Results** We constructed a tissue-level phenome-wide network map of colocalized genes and phenotypes. First, we generated colocalized expression quantitative trait loci from 48 tissues of the Genotype-Tissue Expression project and from publicly available genome-wide association study summary statistics from the UK Biobank. We identified 9,151 colocalized genes for 1,411 phenotypes across 48 tissues. Then, we constructed a bipartite network using the colocalized signals to establish links between genes and phenotypes in each tissue. The majority of links are observed in a single tissue whereas only a few are present in all tissues. Finally, we applied the biLouvain clustering algorithm in each tissue-specific bipartite network to identify co-clusters of non-overlapping genes and phenotypes. The majority of co-clusters contains a small number of genes and phenotypes, and 88.6% of co-clusters are found in only one tissue. To demonstrate functionality of the phenome-wide map, we tested if these co-clusters were enriched with known biological and functional gene classes and observed several significant enrichments. Furthermore, we observed that tissue-specific co-clusters are enriched with reported drug side effects for the corresponding drug target genes in clinical trial data. **Conclusions** The phenome-wide map provides links between genes, phenotypes and tissues across a wide spectrum of biological classes and can yield biological and clinical discoveries. The phenome-wide map is publicly available at https://rstudio-connect.hpc.mssm.edu/biPheMap/.

PrgmNr 2727 - A tissue-specific allelic hierarchy predicts phenotypic outcomes for *USH2A*-related disorders in the RUSH2A study

View session detail

Author Block: **R. B. Hufnagel**[1], B. Guan[2], E. Ullah[1], W. Liang[3], C. C. Brewer[4], J. L. Duncan[5], K. H. Branham[6], S. P. Daiger[7], A. R. Ayala[3], I. Audo[8], J. K. Cheetham[9], T. A. Durham[9], E. Heon[10], C. B. Hoyng[11], A. Iannaccone[12], C. N. Kay[13], M. Michaelides[14], M. E. Pennesi[15], M. S. Singh[16], Foundation Fighting Blindness Consortium Investigator Group; [1]Natl. Eye Inst., Bethesda, MD, [2]NEI/NIH, Bethesda, MD, [3]Jaeb Ctr. for Hlth.Res., Tampa, FL, [4]Natl. Inst. on Deafness and Other Communication Disorders, Bethesda, MD, [5]UCSF, San Francisco, CA, [6]Univ. of Michigan, Kellogg Eye Ctr., Dept. of Ophthalmology and Vision Sci., Ann Arbor, MI, [7]The Univ. of Texas Hlth.Sci. Ctr., Houston, TX, [8]UMR S 968 INSERM, Paris, France, [9]Fndn. Fighting Blindness, Columbia, MD, [10]Hosp for Sick Children, Toronto, ON, Canada, [11]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [12]Duke Univ. Med. Sch., Duke Eye Ctr., Dept. of Ophthalmology, Durham, NC, [13]Vitreoretinal Associates, Gainesville, FL, [14]Moorfields Eye Hosp. and UCL Inst. of Ophthalmology, London, United Kingdom, [15]Casey Eye Inst. - Oregon Hlth.& Sci. Univ., Portland, OR, [16]Wilmer Eye Inst., John Hopkins Univ., Baltimore, MD

Disclosure Block: **R.B. Hufnagel:** None.

Purpose/Introduction: Natural history studies of rare diseases permit genotype-phenotype correlations, combining clinical descriptions and molecular diagnostics. The Rate of Progression in *USH2A* Related Retinal Degeneration (RUSH2A) study evaluated genotype-phenotype correlations among patients with disease-causing variants in *USH2A*.**Materials and Methods:** We assessed molecular genetic findings and the visual, auditory, and olfactory phenotypes of 127 participants with Usher syndrome (USH2) (n=80) or non-syndromic autosomal recessive retinitis pigmentosa (ARRP) (n=47) due to *USH2A* variants.**Results:** *USH2A* truncating alleles were associated with USH2 and had a dose-dependent effect on hearing loss severity with no effect on visual loss severity. A group of missense alleles, in an inter-fibronectin domain, appeared to be hypomorphic for photoreceptor degeneration in ARRP. These alleles were associated with later age of onset, larger visual field area, better sensitivity thresholds, and better electroretinographic responses. No effect of missense variants on auditory or olfactory deficit severity was observed. **Conclusion:** This study expands our knowledge of a tissue-specific *USH2A* allelic hierarchy and adds important prognostic implications for genetic counseling and treatment trial endpoints. Genotype-driven predictions of clinical outcomes may inform clinical trials for allelic disorders with pleiotropic phenotypes.

PrgmNr 2728 - Assisted reproductive technologies reduce fetal growth and alter maternal and fetal DNA methylation

View session detail

Author Block: **D. William**; Ctr. for Fertility and Hlth., Norwegian Inst. of Publ. Hlth.,, Oslo, Norway
Disclosure Block: **D. William:** None.
Numerous observational studies have reported associations between assisted reproductive technologies (ART) and perinatal outcomes. However, the causal nature of these associations remains unclear.
We performed a Mendelian randomization study using a Norwegian cohort of 27,368 genotyped mother-father-newborn trios (613 were ART-conceived children) and investigated the causal effects of ART on birth weight, birth length, maternal and fetal DNA methylation (DNAm). Among these 27368 trios, 745 trios with naturally conceived children and 480 trios with children born via ART have DNA methylation (DNAm) information from Illumina 850K. We estimated the effect of ART using a new unbiased one-sample MR method we have developed called â□□Cross-fitting for Mendelian Randomizationâ□□ (CFMR). We assessed the causality of ART using a powerful pleiotropy-free polygenic risk score as an instrument for ART (area under the curve=0.77 in the test set) using paternally non-transmitted genotypes. We performed an epigenome-wide MR of ART on maternal and fetal DNAm.
We detected 89 CpGs and 158 CpGs causally associated with ART in the children and the mothers, respectively (FDR

PrgmNr 2729 - Enigmas of *TRAF3*: Toward a mechanistic understanding of world differences in gallstone prevalence

View session detail

Author Block: **O. Clay**, S. A. Dingsdag, G. A. Quintero; Univ. del Rosario, BogotÃi, Colombia
Disclosure Block: **O. Clay:** None.
Introduction. An association of the pleiotropic *TRAF3* locus (TNF receptor-associated factor 3) with gallstone prevalence/disease has been observed for a Chilean, largely Mapuche (Native American) population; searches for a similar association in Europe failed (Bustos *et al.*, 2019). Native-ancestry American populations have highest gallstone (and/or type 2 diabetes) prevalence worldwide; if the link is causal, it begs the question of the underlying mechanism(s). Recent attention has been given to potential roles of the bacterial microbiome in gallstone lithogenesis as bacteria are frequently found in the bile and gallstones of patients undergoing cholecystectomy. TRAF3 has roles in innate immune defense against infecting microbes as well as in orchestrating exocyst-mediated extraction of pathogenic bacteria from (bladder) epithelial cells, for removal by immune clearance or body fluids (Miao *et al.*, 2016).
Materials and Methods. Data from 1000 Genomes, including new end-to-end sequences (Ebert *et al*., 2021), were analyzed to identify and characterize the haplotype blocks overlapping and flanking the *TRAF3* gene.
Results. On the basis of a strong CpG island, a corresponding repeat structure varying structurally among haplotype classes, and published methylation data, the promoter region of the 5â gallstone-associated haplotype block of *TRAF3* could be a prime channel allowing common natural variation to influence gallstone prevalence by modulating *TRAF3* transcription. The key structural features (e.g., zinc finger pattern) of the TRAF3 protein and 3â region are encoded by a different haplotype block, located immediately downstream of the promoter blockâs lead SNPs for gallstone association, suggesting independent modulating of the two blocks by common variation. Inter-population differences among the shared haplotypesâ frequencies do not explain why the gallstone effect is observed in Chile but not in Germany.
Conclusion. The salient features of the *TRAF3* promoter haplotype suggest involvement of CpG density-dependent CpG methylation, an epigenetic effect that could be locally modulated, as a likely causal mechanism for the population-specific gallstone association observed and then consolidated by *TRAF3* expression differences in gallbladder and duodenum (Bustos *et al.*, 2019). This partly epigenetic control route could have coevolved with (post)-transcriptional regulation after the Beringian epoch, and with different microbiome backgrounds and dysbioses affecting the biliary tract. A picture emerges of how TRAF3 might act to deter or influence gallstone etiologies through a very different route than other known gallstone genes.

PrgmNr 2730 - Gene expression differences by race and genetic ancestry in B-cell acute lymphoblastic leukemia

View session detail

Author Block: **F. Barragan**[1,2], L. J. Mills[3,4], A. Raduski[2], E. Marcotte[2,4], L. Spector[2,4], L. Williams[2,4], K. Grinde[1]; [1]Macalester Coll., Saint Paul, MN, [2]Div. of Epidemiology & Clinical Res., Dept. of Pediatrics, Univ. of Minnesota, Minneapolis, MN, [3]Dept. of Pediatrics, Univ. of Minnesota, Minneapolis, MN, [4]Masonic Cancer Ctr., Univ. of Minnesota, Minneapolis, MN

Disclosure Block: **F. Barragan:** None.

There are documented differences in clinical presentation and survival by race/ethnicity in children with B-cell acute lymphoblastic leukemia (B-ALL) in the United States: typically, African American (AA) children present with more aggressive cancers than their European American (EA) and Latinx (LAT) counterparts and have worse survival despite EA and LAT children being more frequently diagnosed with B-ALL. Prior work has indicated that racial/ethnic differences in B-ALL outcomes cannot be wholly explained by differences in socioeconomic status or access to care. Related work has also suggested that there may be some genetic factors that could partially explain these residual differences. Using gene expression data generated via microarray and RNA-sequencing on children with B-ALL from the National Cancer Instituteâ□□s TARGET Database (N=324) and St. Judeâ□□s PeCan Server (N=303), we perform differential gene expression analyses by reported race/ethnicity in both datasets, using Limma and DESeq2 with Benjamini-Hochberg (BH) corrections for multiple comparisons. In PeCan using RNAseq data, we identify 374, 198, and 50 statistically significantly differentially expressed (DE; BH adjusted-p

PrgmNr 2731 - Genetic Management in Benign Prostatic Hyperplasia

View session detail

Author Block: J. A. Davalos; Retired, Quito, Ecuador
Disclosure Block: J.A. Davalos: None.
Benign Prostatic Hyperplasia (BPH) is a common disease in elderly. BPH is associated with an increased incidence of hereditary prostate cancer. An updated review was made on genetics in BPH in some papers and publications of several countries such as United States of America, China, Italy, Spain, Argentina and Ecuador. A genetic panel testing could help to identify that risk, but this test hasnâ□□t large diffusion especially in some developing countries. The aim of this review is to present updated information on genetics related to BHP.

PrgmNr 2732 - Identifying casual effects of lipids in neurological diseases using univariate and multivariate Mendelian Randomization

View session detail

Author Block: **Y. Huang**, D. Baird, C-Y. Chen, H. Runz, E. A. Tsai; Biogen, Cambridge, MA
Disclosure Block: **Y. Huang:** Salary/Employment; Biogen.
Lipids metabolism is implicated in neurological diseases and underlying mechanisms suggest modulating cholesterol could be beneficial for neurological disorders, but the effects haven't been thoroughly explored. We used Mendelian Randomization (MR) to evaluate causal effects of lipids in multiple neurological and psychiatric diseases, as well as several brain-related traits. In total, five lipids, HDL, LDL, triglycerides, apolipoprotein A (ApoA), apolipoprotein B (ApoB), and 10 diseases, Alzheimer's (AD), Parkinsonâ□□s, multiple sclerosis (MS), autism, epilepsy, schizophrenia (SCZ), ADHD, bipolar disorder, FTD and major depressive disorder, were included in the analysis. We also studied three brain-related traits, brain volume, education and cognitive performance, to assess causal effects of lipids on intermediate phenotypes. In the univariate analysis, we selected independent genetic instruments (p10), heterogeneity test (Cochranâ□□s Q) followed by radial analysis and visual inspection of the scatter plots. MR was conducted using R package â□□TwoSampleMRâ□□ with five methods: inverse-variance weighting (IVW), MR Egger, weighted median, weighted mode and simple mode. Three protective effects were identified including ApoA for AD (p=2e-9), HDL for education (p=7e-6), and triglycerides for general epilepsy (p=1e-6). Interestingly, our results revealed a protective effect of genetically predicted triglyceride level for epilepsy that is congruent with the ketogenic diet used to control seizures for epilepsy patients. Several suggestive (pHMGCR, *PCSK9* and *LDLR* to explore potential therapeutic effects on neurological traits. We found higher LDLR expression associated with reduced MS risk, but the signal appeared to be driven by a trans-eQTL at the *HLA* locus. MR analyses suggest causal effects of lipids in several neurological disease. Further examination is needed to confirm these findings and understand the underlying mechanisms.

PrgmNr 2733 - Investigating risk factors for perceived facial ageing: A Mendelian randomization study

View session detail

Author Block: **B. Scanlan**, L. Howe, S. Lewis, M. Lyon; Univ. of Bristol, Bristol, United Kingdom
Disclosure Block: **B. Scanlan:** None.

Biological age is one of the most important risk factors for a host of complex diseases because of age-related progressive tissue degeneration that disrupts vital organ architecture and function. At present, there is no gold standard for measuring biological age and existing quantitative techniques are invasive and costly. Perceived age, i.e. how others estimate your age based on your appearance, has been shown to be a reliable predictive biomarker of general cognitive and physical health. For example, looking older than your chronological age has been consistently linked to increased mortality and morbidity. Twin studies suggest a large proportion (approximately 40%) of variation in perceived age is driven by non-genetic factors. Understanding how various biological pathways influence tissue senescence and consequently disease risk and lifespan will identify how to best promote health and longevity in an aging population.

To improve our understanding of factors that influence biological ageing, we performed two-sample Mendelian randomization analyses using Genome-wide Association Study summary statistics to estimate effects of 17 exposures (including BMI, smoking status, alcohol consumption, water intake, educational attainment and age at menarche) on perceived facial ageing. Perceived facial ageing, was measured in UK Biobank (N = 423,999) using self-reported questionnaire data, participants selected whether they are generally told they look â younger than you areâ, â older than you areâ or â about your ageâ. We performed various sensitivity analyses to confirm the robustness of our findings including within-family analyses using sibling pairs.

Mendelian randomization analyses indicated that higher BMI ($P=2.14x10^{-22}$) and smoking heaviness ($P=2.69x10^{-5}$) may lead to being perceived as older than chronological age. Genetic variants associated with skin colour ($P=2.94x110^{-14}$), age at menarche ($P=1.32x10^{-8}$) and male balding patterns ($P=1.37x10^{-26}$, $1.22x10^{-3}$ and $4.78x10^{-16}$ for three balding types) were also associated with perceived facial ageing. Our results provide genetic evidence for effects of physically observable and substance-related phenotypes on perceived facial ageing.

PrgmNr 2734 - Investigating the genetic interplay between adult height and disease

View session detail

Author Block: **A. Papadopoulou**[1], S. Raghavan[2], E. LITKOWSKI[3], M. Graff[4], C. Avery[5], Z. Wang[6], R. A. Smit[7], G. Chittoor[8], N. Josyula[9], W. Zhu[10], J. E. Below[11], S. Vedantam[12], L. Yengo[13], S. I. Berndt[14], J. Arias[15], C-T. Liu[16], A. R. Wood[17], A. Justice[18], Y. Sun[19], J. N. Hirschhorn[20], R. Loos[21], K. E. North[4], P. Deloukas[22], E. Marouli[23], GIANT Consortium[15]; [1]Queen Mary Univ. of London, London, United Kingdom, [2]Div. of BioMed. Informatics and Personalized Med., Aurora, CO, [3]Univ. of Colorado, Aurora, CO, [4]Univ North Carolina, Chapel Hill, NC, [5]Dept. of Epidemiology, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [6]The Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]Icahn Sch. of Med. at Mount Sinai, New York, NY, [8]Dept. of Population Hlth.Sci., Geisinger, Danville, PA, [9]Geisinger Hlth.System, Fremont, CA, [10]Nashville, TN, [11]Vanderbilt Univ Med Ctr., Nashville, TN, [12]Boston Children's Hosp, Sharon, MA, [13]Brisbane, Australia, [14]Natl. Cancer Inst., Rockville, MD, [1516]Boston Univ. SPH, Boston, MA, [17]Coll. of Med. and Hlth., Univ. of Exeter, Exeter, Devon, United Kingdom, [18]Geisinger, Danville, PA, [19]Emory Univ, Atlanta, GA, [20]Boston Children's Hosp., Boston, MA, [21]The Icahn Sch. of Med. at Mount Sinai, New York, NY, [22]Queen Mary Univ. London, London, United Kingdom, [23]Barts and The London Sch. of Med. and Dentistry Queen Mary Univ. of London, London, United Kingdom

Disclosure Block: **A. Papadopoulou:** None.

In the frames of the Genetics of Anthropometric Traits - GIANT - consortium, we conducted a large-scale trans-ethnic meta-analysis for height including up to 4M individuals.

To identify traits and disease related outcomes that are affected by genetically determined height, we performed Phenome Wide Association Studies (PheWAS) analyses using a Polygenic Score (PGS) for increased height as the exposure and 1,421 phenotype codes or â□□phecodesâ□□ as the outcome. The phecodes were based on aggregated ICD-10 codes defining specific diseases or traits. The PGS consists of 6,772 conditionally independent variants derived from a trans-ancestry GWAS meta-analysis for adult height, using 50,000 unrelated and randomly sampled European participants of the UK Biobank as LD reference. We examined health-related outcomes from Hospital Episode Statistics (HES) data in the UK Biobank (N= 311,195) and the Million Veteran Project (MVP) (N=217,225). We meta-analysed the PheWAS results obtained from the two cohorts, using a Bonferroni adjusted significance threshold of pThe meta-analysis yielded 161 significant phecodes (pThe identified signals are being further interrogated through Mendelian Randomisation and mediation analyses. The trans-ethnic analysis described here will be also expanded to ancestry-specific groups.

PrgmNr 2735 - Large-scale genome-wide meta-analyses provide insights for the development of new disease modifying targets for osteoarthritis

View session detail

Author Block: **K. Hatzikotoulas**[1], L. Southam[1], C. G. Boer[2], on behalf of Genetics of Osteoarthritis Consortium; [1]Inst. of Translational Genomics, Helmholtz Zentrum MÃ¼nchen, German Res. Ctr. for Environmental Hlth., Neuherberg, Germany, [2]Dept. of Internal Med., Erasmus MC, Med. Ctr., Rotterdam, Netherlands
Disclosure Block: **K. Hatzikotoulas:** None.
Osteoarthritis is one of the leading causes of disability and pain worldwide, with over 300 million people affected. Currently no curative treatments are available. A detailed understanding of disease aetiopathology and novel drug targets are therefore urgently needed. Here, we conducted the largest genome-wide association study (GWAS) meta-analysis for osteoarthritis to date, across 13 international cohorts stemming from 9 populations, in up to 826,690 individuals (177,517 osteoarthritis patients) of European and East Asian descent. We defined 11 phenotypes encompassing all major sites for osteoarthritis at both weight-bearing and non weight-bearing joints. We also performed a sex specific GWAS meta-analysis for each of the 11 osteoarthritis phenotypes separately and carried out a meta-analysis of early osteoarthritis, defined as age at onset younger than 45 years of age. To identify putative effector genes and causal pathways that represent high-value targets for therapeutics, we integrated supportive information from statistical fine-mapping, expression quantitative trait loci (QTL) colocalization, phenome-wide analyses, animal model data, human musculoskeletal and neuronal phenotype data, plasma protein QTL causal inference analysis and examination of quantitative proteomics and RNA sequencing data for all genes within 1Mb of the osteoarthritis-associated variants in primary patient tissues. We identify 100 unique and independently associated risk variants, 52 of which have not been associated with osteoarthritis before. We report the first thumb, spine, female-specific and early age-at-onset osteoarthritis risk single nucleotide variants (SNVs). We identify 637 genes with at least one line of evidence pointing to a putative effector gene. Gene-set enrichment analyses of these genes, identified bone-, cartilage- and nerve- developmental pathways to be significantly associated (FDRPPARD, *NR3C1*, *VDR*, *MAPK14*, *IGF1R*, *CHST3*). The remaining genes, have market authorisation or are in clinical development for other indications, which represent attractive targets for drug repositioning. In this study, we enhance our understanding of the genetic aetiology of osteoarthritis, shed novel biological insights, and provide a stepping stone for translating genetic associations into osteoarthritis drug development.

PrgmNr 2736 - Linkage analyses confirm chromosome 5q35 as a risk locus in African Ancestry population

View session detail

Author Block: **P. Whitehead Gay**[1], F. RAJABLI[1], L. D. Adams[1], T. Starks[2], M. Jean-Francois[3], The NIA-LOAD Family-Based Study, B. Kunkle[1], E. R. Martin[1], C. Reitz[4], G. S. Byrd[2], M. L. Cuccaro[1], J. M. Vance[1], G. W. Beecham[1], M. A. Pericak-Vance[1]; [1]Univ. of Miami, Miami, FL, [2]Wake Forest Univ., Winston-Salem, NC, [3]Miami, FL, [4]Columbia Univ, New York, NY

Disclosure Block: **P. Whitehead Gay:** None.

Genetic studies of African Ancestry (AA) with late-onset Alzheimer disease (AD) have been limited, despite evidence suggesting an increased risk for AD in AA populations. In this study, we performed whole-genome sequencing (WGS) in multi-generational AA AD families from the Research in African American Alzheimer Disease Initiative (REAAADI) and Late-Onset Alzheimer's Disease Family Study (LOAD) to identify rare causal variants influencing AD through linkage and segregation-based approaches. As part of REAAADI and LOAD, data were generated for 51 families (160 affected and 318 unaffected). We performed a multipoint linkage scan using MERLIN software and genotype data to identify the genetic loci likely to carry risk variants. Following the linkage analysis, WGS data were used to prioritize variants in the consensus regions (HLOD > 3) based on segregation with disease among affected individuals, rarity (MAF 5). The multipoint affected only dominant model identified a suggestive significant signal on Chromosome 5q35 (HLOD 3.20, eight families contributing > 0.58). 1-LOD region is flanking with the locus that was previously reported as a suggestive significant locus in two AA AD genome-wide association studies (Reitz et al. 2013 (p-value=6.9 Ã 10-8) and Kunkle et al. 2020 (p-value=2.6 Ã 10-6)). By conducting the segregation analysis using WGS data, we identified 6 rare variants (MAF 5 that segregated with disease in all affected individuals of the family NC94222 that contributed to the linkage peak with the highest LOD score. Remarkably, a rare variant (rs564891878) in the *DOCK2* gene segregates with the disease in a family with 7 AD individuals (family-specific LOD = 1.72). The variant is intronic, with rare allele frequency in the 1kGP reference data set (MAF = 0.0013) and a moderate CADD score (7.3). In this study, multiplex family linkage analysis confirmed the chromosome 5q35 risk locus previously identified in two AA genome-wide association studies. Our AA population-specific finding shows the importance of diversifying population-level genetic data to better understand the genetic determinants of AD on a global scale. Moreover, identified putative damaging rare variants in multiplex families indicates the critical role of rare variation in AD etiology.

PrgmNr 2737 - Sex-specific genetic modifiers of hemolysis in blood donors

View session detail

Author Block: **F. FANG**[1], K. Hazegh[2], T. Kanias[2,3], G. P. Page[4]; [1]GenOmics, Bioinformatics, and Translational Res. Ctr., RTI Intl., Durham, NC, [2]Vitalant Res. Inst., Denver, CO, [3]Univ. of Colorado3.Dept. of Pathology, Univ. of Colorado, Denver, CO, [4]GenOmics, Bioinformatics, and Translational Res. Ctr., RTI Intl., Atlanta, GA

Disclosure Block: **F. Fang:** None.

Sex differences have been widely observed for many human traits and common diseases. Despite the increasing effort to account for sex in research design and data analysis, the underlying genetic mechanisms leading to such sex differences remains poorly understood. We investigated the sex differences in red blood cell (RBC) predisposition to spontaneous (cold storage) or stress (osmotic fragility, oxidative) hemolysis using data from the NHLBI Recipient Epidemiology Donor Evaluation Study (REDS)-III RBC-Omics cohort of 13,403 multi-racial American blood donors. RBCs undergo increasing hemolysis as they age and when stored after donation. Elevated RBC hemolysis is associated with many adverse outcomes of transfusion. Sex has been known to be related to differences in RBC metabolism and hemolysis in cold storage. These observations have motivated us to identify sex-specific genetic modifiers of hemolysis in blood donors.

We first conducted sex-stratified genome-wide association studies (GWAS) for osmotic hemolysis, which is a measurement of RBC membrane integrity shown in our previous analyses to be modulated by ion channels such as *PIEZO1*, and *ANK1*. Two variants in *SLC4A1* (rs1476512 and rs13306780) reached genome-wide significance (PSLC4A1 encodes for Anion Exchanger-1 (AE1 or band 3) protein, which is the most common protein in the RBC cell membrane. Mutations in *SLC4A1* are associated with hereditary spherocytosis, acidosis, and changes in the Diego blood group. Using the gene expression data from the Genotype-Tissue Expression (GTEx) project, we also identified a significantly higher expression of *SLC4A1* in female whole blood cells (T-test, P=0.039). Male-specific GWAS hits for oxidative hemolysis were found on chromosome 7 surrounds *IMPDH1*, the gene that encodes for the metabolic enzyme inosine monophosphate dehydrogenase. Additionally, we conducted genome-by-sex interaction analysis by incorporating an SNP-by-sex interaction term in GWAS. Several SNPs (rs7739752, rs6922548) surrounding *DEF6* and *PPARD* genes were identified with opposing effects on osmotic hemolysis (i.e., the same SNPs are associated with decreased hemolysis in males and increased hemolysis in females). These findings have led us to the conclusion that sex, as a biological variable, modulates the hemolytic response to genetic mutations that regulate RBC function. Such sex-specific genetic regulation of RBC biology and pathology may further advance the fields of precision transfusion medicine and genetic studies of hemolytic diseases.

PrgmNr 2738 - Unbiased metabolomics by Mendelian randomisation links serum isoleucine to risk of amyotrophic lateral sclerosis

View session detail

Author Block: **J. Cooper-Knock**[1], S. Boddy[1], M. Islam[1], J. Kurz[1], T. Julian[1], C. Harvey[1], T. Moll[1], H. Ghahremani Nezhad[1], J. Marshall[1], K. Kenna[2], S. Zhang[3], M. P. Snyder[4], P. Shaw[1], C. McDermott[1]; [1]Univ. of Sheffield, Sheffield, United Kingdom, [2]UMC Utrecht, Utrech, Netherlands, [3]Stanford Univ., Stanford, CA, [4]Stanford Univ, Stanford, CA
Disclosure Block: **J. Cooper-Knock:** None.
Amyotrophic lateral sclerosis (ALS) is an incurable and rapidly neurodegenerative disease which affects 1/350 individuals. The cause of ALS is unknown but the majority of cases are thought to result from a complex gene-environment interaction. Two sample Mendelian randomisation (MR) enables causal inference between environmental exposures, such as serum metabolite concentrations, and disease risk. We obtained summary statistics from genome-wide association study (GWAS) of serum concentrations of 974 metabolites which were population matched with a GWAS study of ALS. For each metabolite we performed two sample MR using a liberal instrument (p15) instruments as this is likely to lead to false positive results due to instrument pleiotropy. After filtering p-value statistics were not significantly inflated (Î»GC=1.1). After Bonferroni multiple testing correction three metabolites were significantly related to ALS risk: Estrone-3-sulfate (p=6.58e-05, beta=-0.03, se=0.008) and bradykinin (p=8.07e-05, beta=-0.05, se=0.01) were protective which is consistent with literature describing a male preponderance of ALS and the protective effect of medications known to increase serum bradykinin. Serum isoleucine was positively associated with ALS risk (p=1.29e-04, beta=0.05, se=0.01). For each of these three metabolites MR results remained significant using robust measures and the association was confirmed in a distinct ALS GWAS, moreover there was no evidence of genetic pleiotropy or instrument heterogeneity. Isoleucine is metabolised via methylmalonyl-CoA mutase (MCM) in a reaction which consumes vitamin B12. Vitamin B12 insufficiency has been previously associated with ALS risk and our MR analysis revealed that serum vitamin B12 is protective against ALS (p=0.005, beta=-0.17, se=0.06). Multivariate MR revealed that the toxic effect of isoleucine is dependent on vitamin B12 (p>0.05 after correcting for serum vitamin B12). Leucine is a structural isomer of isoleucine but it is not metabolised via MCM; interestingly our MR analysis revealed no relationship between serum leucine and ALS risk. It is notable that a clinical trial of administration of branch chain amino acids, including isoleucine, was previously terminated because of excessive deterioration in the treatment group; this would be consistent with our conclusion that isoleucine exacerbates neuronal damage in ALS via depletion of vitamin B12. We suggest screening of ALS patients for high serum isoleucine levels and supplementation with vitamin B12.

PrgmNr 2739 - Uncovering the role of genomic structural variations associated with risk of osteoporosis through whole genome sequencing

View session detail

Author Block: **K. Su**[1], Y. Liu[1], A. Liu[1], J. Greenbaum[2,1], Z. Xiao[1], C. Xu[3,4], C. Qiu[1,1], L. Zhang[5], Z. Luo[1], Q. Tian[1], L. Wu[1], L-J. Zhao[1], H. Shen[1], H-W. Deng[1]; [1]Ctr. for BioMed. Informatics and Genomics, Tulane Univ., New Orleans, LA, [2]Tulane Univ, New Orleans, LA, [3]4201 W Mem. Rd, OKLAHOMA CITY, OK, [4]Dept. of Biostatistics and Epidemiology, Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma, OK, [5]Soochow Univ., Suzhou, China

Disclosure Block: **K. Su:** None.

Background: Genomic structural variations (SVs), defined as large scale (>50 bp) structural differences/genomic rearrangements, have been shown to have a substantial amount of heritability and are highly polymorphic in the human genome. These SVs contribute to biological speciation and diversification, and also may represent critical sources of phenotypic variability for a variety of complex human diseases. Osteoporosis, a heritable metabolic bone disorder mainly characterized by low bone mineral density (BMD), has previously been shown to be influenced by small scale copy number variations. However, the contribution of large-scale SVs to osteoporosis susceptibility has not yet been explored. In this study, we aimed to characterize the diverse patterns of SVs and investigate the association between SVs and osteoporosis in the human genome.
Methods: Whole genome sequencing (average depth ~22X) analyses were carried out in 5001 subjects from the Louisiana Osteoporosis Study (LOS). BreakDancer was used to detect SVs for genomic sequencing deletions and a gene-based burden test was applied to test the association between the presence of SVs within a given gene and hip BMD. A false discovery rate procedure was applied to correct for multiple comparisons.
Results:We identified 2,959 genes that contain at least one large deletion, including 155 genes that are commonly (in â¥ 50% of samples) affected by at least 50 bp genomic deletions, 1,915 genes affected in more than one but Conclusion: In summary, this study provided a population-based examination of genomic SVs and presented a novel layer of biological factors which may have an important impact on human bone biology. We also highlighted some potential risk genes that may provide new insights into the pathophysiology of osteoporosis.

PrgmNr 2740 - White matter integrity and nicotine dependence in smokers: evaluating vertical and horizontal pleiotropy

View session detail

Author Block: **Z. Ye**; Maryland Psychiatric Res. Ctr., Dept. of Psychiatry, Sch. of Med., Univ. of Maryland, Baltimore, MD
Disclosure Block: **Z. Ye:** None.

Tobacco smoking is an addictive behavior that supports nicotine dependence and an independent risk factor for cancer and other illnesses. Its neurogenetic mechanisms are not fully understood but may act through alterations in cerebral white matter that shows changes consistent with positive and negative cognitive reinforcement mechanisms. We hypothesized that the vertical pleiotropic pathways, where genetic variants influence a trait which in turn influences another trait, link genetic factors, integrity of cerebral white matter (WM) and nicotine addiction. We tested this hypothesis using individual genetic factors, WM integrity measured by fractional anisotropy (FA), and smoking/nicotine dependence in a large epidemiological sample (N=23,624 participants that have genetic, brain imaging and smoking status data available; N=8,830 participants that have genetic, brain imaging and cigarettes per day (CPD) data available) collected by UK Biobank. We performed genome-wide association study (GWAS) to identify previously reported loci associated with smoking and nicotine dependence. We then evaluated two competing vertical pathways: Genes -> WM integrity -> smoking severity versus Genes -> smoking severity -> WM integrity and a horizontal pleiotropy pathway where genetic factors independently affect both smoking severity and WM integrity. The identified variants were located in important susceptibility genes for smoking-induced diseases such as IREB2. Post-hoc analyses showed the genetic predisposition to smoking may act through vertical pleiotropy pathways that maintain nicotine addiction through changes in cerebral white matter.

PrgmNr 2741 - Amplification of CCND1/IGH@ Fusion in Plasma Cell Myeloma - A Case Report and Review of Literatures

View session detail

Author Block: **J. Liu**, J. Gong, J. Rodriguez, G. Uppal, D. Foster, S. Peiper; Thomas Jefferson Univ., Philadelphia, PA
Disclosure Block: **J. Liu:** None.
The t(11;14)(q13;q32) resulting in *CCND1/IGH* fusion is a frequent cytogenetics abnormality in plasma cell myeloma. Overexpression of cyclin D1 dysregulates the normal cell cycle, inhibits differentiation, promotes uncontrolled cellular proliferation and carcinogenic phenotypes.
We describe a 77- year-old female with a diagnosis of plasma cell myeloma and amplification of the *CCND1/IGH* fusion gene. This patient presented with back pain. She had a history of stage II, triple positive breast cancer, osteoporosis, worsening neurological deficits and a right MCA infarct. Innumerable small lucencies throughout chest, abdomen and pelvis along with thoracic and lumbar compression fractures were observed by imaging. Laboratory study showed worsening of known microcytic anemia, low albumin, normal calcium and creatinine. SPEP showed a 1.5g/dl M protein IgA Lambda. Fluorescence in situ hybridization (FISH) panel performed on non-enriched cultured cells demonstrated multiple positive findings suggestive of hyperdiploidy with complex chromosomal abnormalities, including tetrasomy 11, rearrangement of *IGH*, translocation of CCND1/IGH with or without amplification of the fusion product, and deletions of *RB1* and *TP53*. Conventional cytogenetic analysis showed a normal female karyogram most likely due to failed growth of the myeloma cells in culture. Concurrent bone marrow exam showed plasma cell myeloma with 88% bone marrow involvement. Flow cytometric analysis revealed a differential of 23% lymphocytes, 3% monocytes, 65% granulocytes, 24% plasma cells, and Four reports of *CCND1/IGH* fusion amplification have been reported in the literature, to the best of our knowledge, this is the fifth case of *IGH/CCND1* fusion amplification, the third in a patient with plasma cell myeloma. Although difficulty remains to determine the prognostic significance of *CCND1/IGH* fusion amplification due to its rarity, its association with karyotypic complexity and unfavorable outcomes in the published cases suggests a poor prognosis and need of aggressive treatments. Deciphering the underlining molecular pathogenic pathways was attempted. Clinical courses, genetics and laboratory findings and prognostic outcomes were also reviewed and discussed for the five known cases.

PrgmNr 2742 - Cancer Risk C, a functional genomics test, is a sensitive, specific, and rapid diagnostic for Lynch syndrome

View session detail

Author Block: **I. Alim**[1], J. Loke[2], S. Yam[1], S. D. Klugman[3], A. S. Templeton[4], P. A. Newcomb[4], N. M. Lindor[5], R. K. Pai[6], M. A. Jenkins[7], S. Gallinger[8], H. Ostrer[9]; [1]Morgan And Mendel Genomics, Bronx, NY, [2]1300 Morris park Ave Ullmann 817, Bronx, NY, [3]Montefiore Med. Ctr., Bronx, NY, [4]Fred Hutch, Seattle, WA, [5]Mayo Clinic, Scottsdale, AZ, [6]Mayo Clinic, Phoenix, AZ, [7]Univ Melbourne, Carlton, Australia, [8]Univ of Toronto, Toronto, ON, Canada, [9]Albert Einstein Coll. of Med., Bronx, NY

Disclosure Block: **I. Alim:** None.

Heritable mutations in the DNA mismatch repair (MMR) pathway cause Lynch syndrome (LS), a condition that significantly increases risk of colorectal and other cancers. LS mutations are commonly found in the *MLH1*, *MSH2*, *MSH6*, *PMS2* and *EPCAM* genes. Diagnosis of LS is reliant on gene panel sequencing. However, at least half have a variant of uncertain significance (VUS) that cannot be classified for pathogenicity or have no result that informs their diagnosis. Many LS patients are identified only after microsatellite instability testing of tumors is found to be high (MSI-H). We developed a diagnostic test, Cancer Risk C (CR-C), using flow-cytometry based functional variant assays (FVAs) to aid in the diagnosis of LS. Using patient-derived lymphoblastoid cell lines (LCLs) from the Colon Cancer Family Registry (CCFR) an initial cohort was established with either known pathogenic (n=20) or benign variants (n=20). FVAs were tested to identify the effects of pathogenic variants in MMR genes on the nuclear translocation of the MLH1 and MSH2 proteins and the nuclear phosphorylation of the ATM and ATR proteins in response to treatment with the alkylating agent, methylnitronitrosoguanidine. To differentiate pathogenic and benign variants, a risk classification score was developed based on a combination of MLH1, MSH2 and ATR assays. CR-C was 98% sensitive and 95% specific and could be completed within 48 hours. A second cohort of CCFR patient-derived cells with either pathogenic (n=40) or benign variants (n=40) was tested with CR-C and observed to have similar sensitivity and specificity, thus creating a new diagnostic for LS. Amongst those with VUS and MSI-H tumors (n=60), 72% were identified as having LS. This finding contrasted to the previously observed rate of 16% MMR pathogenic gene variants amongst patients with MSI-H tumors. Edited cells with pathogenic variants in MMR genes were rescued from pathogenic to benign when transfected with an expression plasmid containing wild-type cDNA and tested with CR-C. Direct comparison of matched whole blood samples and LCLs yielded comparable results with CR-C (r^2>0.9). Compared to cancer gene panel sequencing, CR-C is more accurate and rapid for diagnosing LS and can be performed on whole blood. When combined with gene rescue experiments in edited cells or LCLs, CR-C can be used to classify VUS as pathogenic or likely pathogenic.

PrgmNr 2743 - Clinically relevant germline variants in a cohort of Mexican young women with breast cancer: analysis from whole-exome sequencing

View session detail

Author Block: **L. GÃ³mez Flores Ramos**[1], M. Dean[2], K. Jones[3], M. Wang[3], C. Villarreal-Garza[4], R. Ã□lvarez-GÃ³mez[5], T. Wegman-Ostrosky[5], N. Reynoso-NoverÃ³n[5], V. Fragoso-Ontiveros[6], L. SÃinchez-Zamorano[7], M. Trujillo MartÃnez[8], L. Grimaldo[1], M. P. Gallegos-Arreola[9], P. Ostrosky-Wegman[10], A. Mohar-Betancourt[5]; [1]CONACYT-Natl. Inst. of Publ. Hlth., Cuernavaca, Mexico, [2]NCI, Rockville, MD, [3]Natl. Cancer Inst., Gaithersburg, MD, [4]5Breast Cancer Ctr., Hosp. Zambrano Hellion, Tecnologico de Monterrey, Monterrey, Mexico, [5]Natl. Cancer Inst., Mexico City, Mexico, [6]Natl. Cancer Inst., Monterrey, Mexico, [7]Natl. Inst. of Publ. Hlth., Cuernavaca, Mexico, [8]Gen. Hosp. No.7-Mexican Inst. of Social Security, Cuautla, Mexico, [9]CIBO, IMSS, Guadalajara, Mexico, [10]14Inst. of BioMed. Res., Natl. Autonomous Univ. of Mexico, Mexico City, Mexico

Disclosure Block: **L. GÃ³mez Flores Ramos:** None.

Young women with breast cancer represent 15% of cancer cases in Latin America. Genomic studies have found that early-onset breast-cancer cases exhibit a higher genetic susceptibility and a specific genomic signature as compared to their older counterparts. The aim of this study was to describe clinically relevant germline variants in a cohort of young women with breast cancer. To achieve this, we analyzed hereditary-cancer genes from whole-exome sequencing (WES) data in 108 unrelated women with an extreme phenotype of breast cancer (â□¤40 years of age), diagnosed and treated at the National Cancer Institute of Mexico (NCIMx). We obtained approval from the Research Ethics Committee of the NCIMx (CEI/1123/16) and provided patients with genetic counseling and written consent. Exome capture and sequencing were conducted at the Laboratory of Translational Genomics, at the NCI, Maryland, USA. The GRCh37 assembly was used as the genome reference. Variants that failed to pass our pipeline quality control metric (CScorefilter), had a read depth â□□0.8 were excluded from the analysis. Variants were filtered using popmaxfreq BRCA2 represented half of the pathogenic variants found in this group and accounted for 6% of hereditary-cancer cases. *BRCA*1 represented 23% of mutations, followed by *PALB2* (15%)*; TP53* and *RAD51C* (8% each). We describe for the first time, in a patient with breast cancer, a pathogenic duplication in *RAD51C* (c.519dupT). The median age at diagnosis was 35 years overall; however, it was six years younger in patients with pathogenic variants. Age at diagnosis (OR=0.82, CI 95% 0.71-0.94; P= 0.008) and first-degree family history of cancer (OR=8.26, CI95% 1.35-50; P= 0.022) were the only epidemiological variables associated with mutational status. We found no differences in DFS (p=0.403) or OS (p=0.735) among mutational status subgroups. Genetic ancestry and tumor subtype were not associated with mutational status.

PrgmNr 2745 - Facilitators and barriers to implementing non-indication based genetic testing within corporate wellness programs

View session detail

Author Block: **J. M. Goehringer**[1], Z. Salvati[1], P. Deverka[2], M. S. Williams[1]; [1]Geisinger, Danville, PA, [2]Deverka Consulting, Raleigh-Durham-Chapel Hill, NC
Disclosure Block: **J.M. Goehringer:** None.
In the era of precision medicine, population health screening for inherited disease risks has gained momentum. Institutions and genetic testing vendors are increasingly promoting preventative practices and risk mitigation strategies, and employers have begun offering genetic testing as an employee health benefit. Stakeholders Assessing Genetics with Employers (SAGE) was funded by NHGRI to evaluate the broad experience of offering non-indication based genetic testing (NIBGT) through employer-sponsored health benefits. NIBGT is defined as voluntary, health-related genetic testing for individuals with or without a personal or family history of genetic disease.
Company leaders of self-insured employers, genetic testing vendors, and researchers who evaluate wellness programs were approached for participation in a qualitative research study. Nine individuals were enrolled and interviewed using a semi-structured interview guide. Episodic summaries, memos, and transcripts were analyzed with thematic coding, guided by the Framework Method. Here we focus on the facilitators and barriers to NIBGT implementation.

Barriers

Implementation of genetic testing in the wellness setting is more challenging than traditional wellness offerings and the evidence to support use as a screening tool is weak. Companies want to avoid being early adopters of genetic testing within the workplace as the return on investment is not yet demonstrated. This would require pooling large amounts data across employers to study outcomes which is hindered by companies avoiding collaboration with competitors. Further, the timeline for observing favorable health outcomes in genetic disease requires extensive longitudinal follow-up. There are perceptions that human resources and insurance carriers add complexity to the implementation process. Provider education is another barrier due to the complexity of genetics and provider concerns regarding managing employee genetic test results. Last, privacy concerns regarding the use of genetic information is prevalent. There is a lack of understanding about genetic information-related legal protections (mainly GINA) among both employers and employees, leading to reluctance to adopting NIBGT.

Facilitators

Approaching the right individuals within a company drives successful implementation of NIBGT within wellness programs. Implementation is often initiated when C-suite executives champion genetic testing and lead decision-making processes. Program implementation can be facilitated with the help of knowledgeable experts from inside the company including geneticists, other physicians, and genetic counselors.

PrgmNr 2746 - Genetic testing in patients with colorectal cancer; a real-world analysis of utilization and results

View session detail

Author Block: **C. Moretz**[1], S. Nielsen[2], J. Dalton[3], K. E. Hatchell[1], S. T. Michalski[1], S. Stacey Dacosta-Byfield[3], J. John Buresh[3], T. Hall[4], R. L. Nussbaum[1], E. D. Esplin[1]; [1]Invitae, San Francisco, CA, [2]Invitae, Chicago, IL, [3]UnitedHlth.Group, Minnetonka, MN, [4]UnitedHlth.Group Res. and Dev., Minnetonka, MN
Disclosure Block: **C. Moretz:** Major Stockholder/Ownership Interest; Invitae.

Background

Colorectal cancer (CRC) affects approximately 104,000 patients (pts) annually in the United States (U.S.), and up to one-third of cases are estimated to be genetic and/or familial. In 2020, a large U.S. insurer established Medical Policy allowing for and reimbursing germline genetic testing (GGT) for all CRC pts. This study reports overall uptake of GGT in CRC pts under this coverage policy with actionable findings and management implications for pts tested. **Methods** Two independent de-identified datasets were reviewed, including administrative claims data of commercially insured (COM) and Medicare Advantage (MA) enrollees from a large national health plan in the U.S., which included pts aged 18+ with CRC (â¥1 claim with ICD10 C18, C19 or C20) who were continuously enrolled (CE) in the health plan from 1/2019-12/2020. GGT was based on CPT codes during 2020. A second dataset of CRC pts whose GGT was performed by a large genetic testing laboratory and was billed to the insurer under the Medical Policy in 2020, was also reviewed. Patient demographics, clinical information and GGT results were descriptively analyzed. **Results** Of the >14 million CE enrollees, 39,382 were newly diagnosed CRC pts in 2020; 16,539 were COM and 22,843 were MA pts. Overall, 5.6% (2,196) of the cohort received GGT (8.1% of COM pts compared to 3.8% of MA pts). Those that received GGT were younger than the overall cohort of CRC pts, with a higher proportion in the COM population. From the GGT dataset, 787 pts had test results available for review. 142 (18%) pts had 152 pathogenic/likely pathogenic (P/LP) variants in 30 genes, including: *MSH2*, *MLH1*, *PMS2*, *MSH6*, *CHEK2*, *APC, BRCA2, ATM, MUTYH* (biallelic). Overall, 130 of the 142 (92%) had P/LP variants in genes with precision therapy, clinical trial and/or published management guidelines. In a subset of pts (n=674) with ethnicity data, Asian, Black/African-American and Hispanic pts showed lower uptake of germline testing relative to Caucasians. **Conclusions** Despite Medical Policy allowing for GGT for all pts with CRC,

PrgmNr 2747 - Prevalence of pathogenic mutations and variants of uncertain significance in Cambridge, MA

View session detail

Author Block: **C. JANI**[1], O. Al omari[1], H. Singh[2], K. Stafford[1], L. Agarwal[1], A. Rupal[1], A. Dar[1], M. Lotz[1], A. Dangelo[1], P. Lam[1]; [1]Mount Auburn Hosp., Cambridge, MA, [2]Med. Coll. of Wisconsin, milwaukee, WI
Disclosure Block: **C. Jani:** None.
Objective Genetic testing allows for enhanced prognostication and early intervention in patients with a high risk of developing cancer. The most common inheritable cancer syndromes include Hereditary Breast and Ovarian Cancer (HBOC) and Lynch Syndrome (LS). Genetic testing often reveals variants of uncertain significance (VUS), for which association with disease risk is unclear. The ambiguity of this finding creates a dilemma for everyone. In this retrospective observational study, we seek to determine the prevalence of pathogenic mutations and VUS identified in patients undergoing genetic testing in a community hospital in Cambridge, MA. **Methods** We included patients undergoing genetic testing at our hospital between July 2018 and October 2020. Medical charts were abstracted to identify patient demographics, clinical characteristics, family cancer history, indication for genetic testing, mutations and VUS identified, and subsequent management. As our study period spanned during the COVID-19 pandemic, we also assessed its impact on care accessibility. **Results** We reviewed data for 663 patients. The population had a mean age of 50 years (S.D 15). 597(90%) were females, 63 (9.5%) were males and 2 were transgender. There was a high proportion of Ashkenazi Jewish descent (14.3%). First-degree family had cancer history in 530 (79.9%) patients whereas 552 (83.3%) had positive cancer history in the second-degree family. 265 (39.9%) had a personal history of cancer, most commonly breast (214). The most common indication for genetic testing was HBOC (558, 84.2%). Out of 76 (11.4%) deleterious mutations, the most common were BRCA mutations (23, 30%) followed by MUTYH (10,13.1%), and Lynch mutations (5, 6.5%). Out of 200 (30.1%) VUS, the most common were APC (23). We found no significant trend in genetic counseling consultations over our 11 months (12/2021-10/2020) of COVID-19 part of our study period despite the pandemic (R^2 = 0.006). **Conclusions:** In our study, prevalence rates of pathogenic mutations and VUS were 11.4% and 30.1% respectively. Despite advances in identifying genetic risk for colon and gynecological cancers, the majority of patients presented for breast cancer-related screening. Due to the high prevalence of VUS, and the challenges faced in addressing these findings, future work needs to explore evidence-based approaches to communication and interpretation of these findings. Lastly, we noted high efficacy in our conversion of in-person genetics consultations to telemedicine during the COVID-19 pandemic, suggesting its robust format.

PrgmNr 2748 - Tumor genomic profiling (TGP) - identified mutations in moderate risk breast and ovarian cancer genes: worthy targets for germline confirmatory testing

View session detail

Author Block: **H. Llorin**[1], M. Graf[2], N. Chun[2], J. Ford[2]; [1]Mass Gen. Brigham, Boston, MA, [2]Stanford Univ., Stanford, CA
Disclosure Block: **H. Llorin:** None.
Tumor genomic profiling (TGP) has become more widespread and clinically actionable through advancements in sequencing technology and molecular oncology. Recent changes in oncology practice guidelines indicate that mutations in cancer susceptibility genes identified on TGP should prompt confirmatory germline testing; however, there are no established best practices for identifying which somatic mutations are most appropriate for germline confirmation. Our study aimed to determine the proportion of patients with TGP-identified mutations in moderate risk breast and ovarian cancer genes (ATM, BRIP1, CHEK2, PALB2, RAD51C, and RAD51D) who previously would not have been considered for germline testing based on previous clinical and family history. We retrospectively analyzed adult Stanford Health Care patients with any type of advanced cancer who had undergone solid tumor genomic profiling. From 7468 tumor genomic profiling reports, we identified 166 patients with TGP-identified mutations in moderate risk breast and ovarian cancer susceptibility. Retrospective chart reviews were performed on 160 patients on whom we had adequate information to determine demographic characteristics, cancer history, eligibility for germline testing, and germline testing results. Nearly half (45.3% [73/160]) of patients would not have been eligible for germline testing if not for a somatic mutation in the gene of interest. Of the 64 cases that underwent multigene panel germline testing that included our genes of interest, about half (51.5% [33/64]) had results that confirmed germline origin of the somatic finding. High rates of germline confirmation were found in PALB2 (100% [5/5]), ATM (40% [14/35]), CHEK2 (61.5% [8/13]), and BRIP1 (57.1% [4/7]). Furthermore, 13.7% (n=22) of cases had a founder mutation on TGP. This data shows that recent guideline changes increase eligibility for germline testing among patients with somatic mutations in moderate risk breast and ovarian cancer susceptibility genes. Traditional germline testing guidelines based on personal and family history do not adequately capture hereditary cancer syndromes among a cohort of advanced cancer patients. Rates of germline confirmation are ~40% or higher in patients with TGP-identified mutations in ATM, CHEK2, and PALB2. In the context of limited healthcare resources, more study is needed to determine which TGP-identified mutations should be prioritized for referral to cancer genetics to maximize hereditary cancer syndrome diagnoses.

PrgmNr 2749 - Disease risk prediction with two-phase study data

View session detail

Author Block: **D. Soave**[1], J. Lawless[2], N. Cheng[3], P. Awadalla[4]; [1]Wilfrid Laurier Univ., Waterloo, ON, Canada, [2]Univ. of Waterloo, Waterloo, ON, Canada, [3]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [4]Ontario Inst Cancer Res., Toronto, ON, Canada
Disclosure Block: **D. Soave:** None.
The prognosis of many cancers is significantly improved when detected at early stages. Through retrospective selection of two-phase (2P) study samples, researchers can leverage the prospective nature of population health cohorts to investigate the relationships between expensive molecular information and rare diseases, such as cancer, prior to traditional clinical diagnosis. When carefully selected, 2P studies provide an efficient and cost-effective alternative to designs requiring the complete cohort. Outcome and covariate dependent sampling designs are needed for a 2P study, and proper specification of design weights is critical for aspects of both the building and assessment risk models. Often, however, the sampling mechanisms are misspecified or altogether ignored when analyzing the data. Here, we investigate techniques and tools for assessment of risk prediction model using 2P data. Unbiased assessment of model performance at the population level is estimated through pseudo (weighted) metrics of discrimination and calibration. We present the theoretical and numerical properties of weight-adjusted estimates of concordance, Kullback-Leibler score, Brier score, and time-dependent receiver operating characteristic curves. We further investigate the specification of design weights and baseline hazard estimation and the use of weight-adjusted criteria in the context of resampling based strategies for model selection. With the goal of detecting cancer prior to the onset of symptoms (and prior to diagnosis through clinical screening), we profiled cfDNA methylation in blood samples from pre-cancer (collected up to 7 years prior to the detection of breast or prostate cancer) and matched control participants (no cancer on follow-up under study) retrospectively selected from the Canadian Partnership for Tomorrowâ□□s Health (CanPaTH) study. Using a novel implementation of weighted regularized regression for two-phase failure time data, we develop and assess cancer risk prediction models using cfDNA methylation markers to predict who will and will not develop cancer up to 7 years prior to clinical detection.

PrgmNr 2750 - Evidence of divergent genetic associations across breast cancer molecular subtypes via distal mediator-enriched Transcriptome-Wide Association Study

View session detail

Author Block: **A. Patel**[1], M. Troester[1], M. Love[2], A. Bhattacharya[3]; [1]Univ. of North Carolina - Chapel Hill, Chapel Hill, NC, [2]UNC-Chapel Hill, Chapel Hill, NC, [3]Univ. of California at Los Angeles, Carrboro, NC
Disclosure Block: **A. Patel:** None.
Background: Breast cancer (BC) molecular subtype (Luminal A-like, Luminal B-like, HER2, Basal-like) is a central prognostic factor for BC mortality. Of the 210 genetic variants associated with risk for BC, many show differential associations across subtype. These variants are predominantly in non-coding regions, and genes and cellular contexts giving rise to the risk from these variants remain largely unknown. Approaches such as Transcriptome-Wide Association Study (TWAS), which offer gene-level assessment of association of variants, can bridge this gap and allow discovery of genetic mechanisms for known and novel loci. **Objective**: We aimed to identify genes whose germline-regulated expression is associated with risk of BC subtypes across healthy breast, adipose, fibroblast, and immune tissue. **Methods**: Using GTEx V8, we constructed predictive models of gene expression from germline variation in European ancestry individuals. Importantly, predictive models leveraged MOSTWAS and were enriched for distal regulatory variants in addition to local (1 Mb surrounding gene) variation. Genes with nominally significant h^2 (p R^2 for predicted and observed expression > 0.01 with p p = 3.4 x 10^{-7} (0.05/146,608 gene-tissue-subtype pairs) and we followed-up on significant genes with a permutation test on predictive model weights (FDR-adjusted p Results: At TWAS-significance and permutation threshold, we identified 353, 229, 184, and 203 genes associated with Luminal A-like, Luminal B-like, HER2, and Basal-like BC, respectively, across etiologically relevant tissue. For healthy breast tissue, 92, 49, 41, and 46 genes, respectively, were identified. Many identified genes were at loci not uncovered in GWAS. Across etiologically relevant tissues, only two genes (*CBX8*, *STK19*) were associated across all subtypes. For healthy breast tissue, only *CBX8*, a known oncogene, was associated across all subtypes, and associations were consistent in direction but differing in magnitude. **Conclusion**: Our findings suggest marked differences in the germline genetic basis of BC subtypes, which may be relevant for developing subtype-specific risk prediction models. Findings also suggest a high level of tissue-specificity for putative genes underlying etiologic heterogeneity for BC.

PrgmNr 2751 - Genetic susceptibility to smoking and alcohol use behaviors identifies differential risk in oropharyngeal and oral cavity cancer patients based on HPV status

View session detail

Author Block: **A. Thakral**[1], J. Lee[2,1], S. Virani[3], A. Olshan[4], A. Ness[5,6], N. Hayes[7], T. Waterboer[8], B. Diergaarde[9], P. Brennan[3], R. Hung[10], G. Liu[1,2], W. Xu[1], O. Espin-Garcia[11]; [1]Princess Margaret Cancer Ctr., Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Intl. Agency for Res. on Cancer, Genomic Epidemiology Branch, Lyon, France, [4]Univ. of North Carolina Gillings Sch. of Global Publ. Hlth.,, Chapel Hill, NC, [5]Bristol Dental Sch., Univ. of Bristol, Bristol, United Kingdom, [6]Natl. Inst. for Hlth.Res. (NIHR) Bristol BioMed. Res. Ctr., Univ. Hosp. Bristol and Weston NHS Fndn. Trust and the Univ. of Bristol, Bristol, United Kingdom, [7]Div. of Med. Oncology and Ctr. for Cancer Res., Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [8]Div. of Infections and Cancer Epidemiology, German Cancer Res. Ctr., Heidelberg, Germany, [9]Univ of Pittsburgh, Pittsburgh, PA, [10]Lunenfeld-Tanenbaum Res. Inst., Sinai Hlth.System, Toronto, ON, Canada, [11]Princess Margaret Cancer Ctr., Toronto, Canada

Disclosure Block: **A. Thakral:** None.

Background: A recent meta-GWAS conducted by the GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN) has identified several loci associated with tobacco and alcohol use behaviors. We investigated how these findings translate in head and neck cancer (HNC). **Aim**: To investigate the association of genetic susceptibility to tobacco and alcohol use with risk of HNC within the context of Human Papillomavirus (HPV) infection. **Methods**: *Study population*: European ancestry individuals participating in the VOYAGER consortium, an international multi-site study focusing on the impact of HPV serology, germline, and tumor genomics on risk and outcome of HNC: 3580 cases of oral cavity (OC)/oropharyngeal cancer (OPC), 3240 healthy controls. *Classification:* HPV serology and p16 tumor status were used to classify cases based on HPV status: 1143 HPV-positive OPC, 2437 HPV-negative HNC [JL1] (HPV-negative OPC/OC). *Genotype data* were obtained from the VOYAGER consortium using the OncoArray chip. *Imputation* was performed on TOPMed Imputation Server using the TOPMed r2 panel. *Polygenic Risk scores* (PRS) were computed using LDpred2 and GSCAN summary statistics for smoking status, age of smoking initiation, cigarettes per day, smoking cessation, and number of drinks per week. *Analysis*: Firstly, measured behaviors were regressed using mixed effects models on their respective PRS with following covariates: age, sex, age-sex interaction, first 7 genetic principal components, and a random intercept for study site. Next, the outcome (HPV-positive OPC, HPV-negative HNC, control) was regressed, via a mixed effects logistic model, on each PRS, with the same covariates and random intercept as the first analysis. Finally, mediation analyses tested the proportion of effect of each PRS on both cancer groups, mediated through measured behavior. **Results**: Measured behaviors were significantly associated with their corresponding PRS across all populations (HPV-positive OPC, HPV-negative HNC, controls). Only PRS of smoking behaviors were associated with risk of HPV-positive OPC, while PRS of both smoking and alcohol behaviors were associated with risk of HPV-negative HNC. These associations were partly mediated through their measured behavior. Mediated effects were significantly larger for HPV-negative HNC than HPV-positive OPC. **Conclusions**: Genetic susceptibility to tobacco use is associated with risk of HPV-positive OPC. In contrast, the genetic susceptibility to both tobacco and alcohol use is associated with risk of HPV-negative HNC. These associations are partly mediated through measured behavior and to a larger extent for HPV-negative HNC than HPV-positive OPC.

PrgmNr 2752 - High prevalence of subclinical micro-FAFs in facial skin in Tuberous Sclerosis Complex

View session detail

Author Block: **K. Klonowska**[1], J. Grevelink[2], K. Giannikou[1], B. Ogorek[1], D. J. Kwiatkowski[1]; [1]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [2]Boston Dermatology and Laser Ctr., Massachusetts Gen. Hosp., Boston, MA

Disclosure Block: **K. Klonowska:** None.

Introduction: Tuberous Sclerosis Complex (TSC) is anautosomal dominant syndrome due to loss-of-function *TSC1*/*TSC2*mutations, characterized by the presence of distinctive hamartomatous tumors inmultiple tissues and organs, including brain, heart, kidney, lungs and skin. TSCtumors develop through a two-hit mechanism, and we had previously shown that UV-inducedmutation was the cause of second hits in TSC facial angiofibromas (FAFs).**Methods:** Recently, we developed a novel MPS method, MultiplexHigh-sensitivity PCR Assay (MHPA), that has a sensitivity of 0.01-0.05% variantallele frequency (VAF) in the most commonly mutated regions of *TSC2*.**Results:** MHPA analysis of 24 FAFs, including 19 from TSC patients with low level systemicmosaicism and 5 with heterozygous germline *TSC2* mutations, led tothe identification of 99 low VAF(0.01-8.02%, median:0.08%,) somatic *TSC2* indels/point mutations. Remarkably,two or more somatic *TSC2* mutations were identified in 19 of 24(79%) FAFs,suggesting that these small (2mm diameter) FAF biopsies contained at least twoangiofibroma clones. 34 of 99(34%)mutations were CC:GG>TT:AA dinucleotide variants (DNVs), indicativeof UV radiation causation, affecting 30 of 522(6%) CC:GG sites in the region sequenced. We alsodeveloped an MHPA assay for *TP53*, and identified an average of 8.1 *TP53* mutations per FAF biopsy (likely occurring in keratinocytes inthese biopsies). The nonsynonymous to synonymous mutation ratio [*TSC2*:10.4(83/8); *TP53*: 29.5(177/6)] in these samples indicatesthat mutations in both genes were likely conferring a selective growthadvantage. Single nucleotide variants (SNVs)/DNVs in *TSC2*/*TP53*were predominantly C:G>T:Aand CC:GG>TT:AA, with more C/CC than G/GG mutated, reflectingtranscription-coupled nucleotide excision repair occurring for UV mutations (untranscribedstrand bias, p=0.0007 and p=0.009, respectively). Our MHPA analysis also led tothe identification of a novel UV-related indel mutation signature, includingcomplex indel-SNV/DNV pattern, in 18 instances.**Conclusions:** Considering the number of mutations, themosaic VAF in FAFs, andthe size of the total facial skin, we estimate that 100-200,000 polyclonalfibroblast proliferations due to second hit mutations in *TSC2* (subclinicalâ□□micro-FAFsâ□□) occur in the skin of TSC patients, a small proportion of whichdevelop into observable FAF lesions. These observations highlight theimportance of sunblock and other measures to limit facial UV exposure for TSCindividuals of all ages to prevent FAF development. *The study was funded by FY2020 Tuberous SclerosisAlliance Postdoctoral Fellowship Award (KK) and Engles Family Fund*

PrgmNr 2753 - PhenX presents a Variable Compare tool for COVID-19 and a new research domain: Cancer Outcomes and Survivorship

View session detail

Author Block: **M. Krzyzanowski**[1], D. Williams[1], L. Cox[1], C. Hill[1], M. Phillips[1], J. Schoden[1], L. Nekhlyudov[2], C. A. McCarty[3], W. K. Chung[4], K. B. Brothers[5], M. D. Ritchie[6], S. Hwang[1], H. Pan[1], J. Waterfield[1], I. Glaze[1], M. Engle[1], S. Armson[1], T. Reeve[1], I. Terry[7], C. Ives[1], N. Braswell[1], S. R. Mitchell[1], G. Swigart[1], A. Riley[1], J. Dayal[8], M. Goldrich[8], E. M. Ramos[9], T. Hendershot[1], C. M. Hamilton[1]; [1]RTI Intl., Research Triangle Park, NC, [2]Brigham and Women's Hosp., Boston, MA, [3]Univ. of Minnesota Med. Sch., Duluth, MN, [4]Columbia Univ, New York, NY, [5]Univ. of Louisville, Louisville, KY, [6]Univ. of Pennsylvania, Philadelphia, PA, [7]LunaPBC, San Diego, CA, [8]NHGRI, Bethesda, MD, [9]NIH, Bethesda, MD
Disclosure Block: **M. Krzyzanowski:** None.
The PhenX (**Phen**otypes and e**x**posures) Toolkit (https://www.phenxtoolkit.org/) is a web-based catalog of recommended measurement protocols and associated bioinformatics tools to assist with study design and facilitate cross-study data integration and analyses. PhenX protocols are recommended by domain experts using a consensus-based process which includes community input. The PhenX Toolkit currently includes more than 800 measurement protocols from 29 research domains (e.g., Demographics and Diabetes) and six collections that add depth to the Toolkit (e.g., Social Determinants of Health) and has over 3,800 registered users. The Toolkit has recently expanded with collections and tools for COVID-19 research. With input from crowdsourcing, 20 COVID-19 protocols were organized into six COVID-19 collections: Behaviors and Risks; Ethnicity, Race and Demographics; History, Treatment, and Outcomes; Information Resources; Psychosocial and Mental Health; and Socioeconomic. Because health disparities were identified as the number one concern for COVID-19 research, PhenX recommends use of the Social Determinants of Health: Core collection. The PhenX COVID-19 Library aggregates COVID-19 protocols in use and includes keyword search to filter protocols. To date, there are 118 collection tools and 517 modules that are annotated with sub-topics such as mental health and avoidant behaviors and main topics that conceptually link modules to the NIH Public Health Emergency and Disaster Research Response (DR2) site. A variable compare tool allows researchers to choose between COVID questionnaires from the COVID-19 collections and COVID-19 Library. This tool includes keyword search at the variable level, side by side comparisons of questionnaires and visual representation of the number of similarities between protocols. PhenX has also added a new domain of protocols for Cancer Outcomes and Survivorship research to complement the existing Cancer domain's focus on risk factors. It includes 15 protocols that are relevant for research focusing on cancer survivors and caregivers. Protocols address a wide range of topics including emotional, physical, and social impact and burden of cancer along with experiences with cancer related care. Also, a working group of experts completed the consensus process and recommended 15 measurement protocols for the new Genomic Medicine Implementation domain. The measurement protocols addressed the following scope elements: knowledge and education (both patients and providers), implementation science, changes in management and treatment, return of results, patient outcomes, and ELSI (ethical, legal, and social issues).

PrgmNr 2754 - Prevalence of autoimmune diseases in Li-Fraumeni Syndrome

View session detail

Author Block: **I. Obregon**, K. de Andrade, P. P. Khincha, R. Bremer, S. A. Savage; Clinical Genetics Branch, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, Rockville, MD
Disclosure Block: **I. Obregon:** None.
Li-Fraumeni syndrome (LFS) is an inherited cancer predisposition syndrome with extremely high cancer risks throughout the lifespan, caused primarily by germline pathogenic *TP53* variants. The effect of germline *TP53* variation on immune function or on immune surveillance in LFS-associated carcinogenesis is not known. In this pilot study, we evaluated the prevalence of autoimmune diseases in individuals with LFS compared to controls. Participants enrolled in the National Cancer Instituteâ□□s longitudinal LFS study (NCT01443468) completed questionnaires inquiring about 24 autoimmune disease diagnoses, age at diagnosis, and need for treatment. Individuals with confirmed LFS-associated *TP53* variants were compared to their non-carrier relatives (controls). Studentâ□□s T-test was used to compare ages between groups. Chi-Square test of independence was performed to assess the relationship between *TP53* variant status and the prevalence of autoimmune diseases. Of the 185 respondents, 113 were carriers of germline *TP53* variants and 72 were controls. There were 82 (72%) female carriers and 44 (61%) female controls. Variant carriers were younger than controls (median age 44 years *vs* 59 years respectively, pvs 46 years respectively, pTP53 variant status and occurrence of autoimmune disease. The predilection of female diagnoses and younger age at onset in the variant carriers may be due to reporting bias, or a true reflection of disease prevalence in this cohort. Further studies of larger cohorts and of *TP53* immunobiology are necessary to validate these results and to understand the role(s) of immune surveillance in cancer in LFS.

PrgmNr 2755 - Prostate cancer susceptibility variants and their interplay with lifestyle factors on the risk of lethal disease

View session detail

Author Block: **A. Plym**[1,2,3], Y. Zhang[2], K. Stopsack[2,4], B. Delcoigne[3], F. Wiklund[3], C. A. Haiman[5], A. S. Kibel[1], E. Giovannucci[2], K. Penney[1,2], L. A. Mucci[2]; [1]Brigham and Women's Hosp., Boston, MA, [2]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [3]Karolinska Inst.t, Stockholm, Sweden, [4]Mem. Sloan Kettering Cancer Ctr., New York, NY, [5]Univ. of Southern California, Los Angeles, CA
Disclosure Block: **A. Plym:** None.
Background While over 260 germline prostate cancer susceptibility variants have been identified to date, few variants have been specifically associated with advanced prostate cancer. We sought to identify variants associated with lethal prostate cancer and their possible interplay with lifestyle factors.
Methods We included 12,411 genotyped men in the Health Professionals Follow-up Study (HPFS) and the Physiciansâ Health Study (PHS) who have been prospectively followed for incident prostate cancer and lethal disease (metastases/cancer-specific death) between 1993 and 2019 (HPFS) and between 1983 and 2010 (PHS). Using Cox regression, we examined the association between 246 previously identified common (0.02 P **Results** 435 men in the cohorts developed lethal prostate cancer. Of the 246 prostate cancer risk variants, 10 were associated with lethal prostate cancer at *P* P = 4.11e-06) for one risk allele increase. The other risk loci included rs4901313 on 14q22 (HR, 1.50; 95% CI, 1.23-1.83; *P* = 5.59e-05), rs150184171 on 6q23 (HR, 1.3; 95% CI, 1.15-1.52; *P* = 1.04e-04), and rs7679673 on 4q24 (HR, 1.33; 95% CI, 1.14-1.55; *P* = 2.19e-04). For two of the variants, rs10090154 and rs7679673, we observed effect modification by lifestyle factors, with the highest risk among men with an unhealthy lifestyle.
Conclusion Our study demonstrates strong associations between previously identified prostate cancer susceptibility variants and lethal disease. We observed a possible interplay with lifestyle factors, which if replicated can help understand the etiology of aggressive prostate cancer.

PrgmNr 2756 - Quantification of cancer risk, patterns, and genotype-phenotype associations in Li-Fraumeni syndrome

View session detail

Author Block: **K. De Andrade**[1], P. Khincha[1], J. Hatton[1], M. Frone[1], T. Wegman-Ostrosky[2], K. Jones[3], K. Teshome[3], P. L. Mai[4], A. Best[5], S. A. Savage[1]; [1]Clinical Genetics Branch, Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., NIH, Rockville, MD, [2]Basic Res. Subdirection, Inst. Natl. de CancerologÃa (INCan), Mexico City, Mexico, [3]Cancer Genomics Res. Lab., Frederick Natl. Lab. for Cancer Res., Frederick, MD, [4]Magee Womens Hosp., Pittsburgh, PA, [5]Biometrics Res. Program, Div. of Cancer Treatment and Diagnosis, Natl. Cancer Inst., NIH, Rockville, MD

Disclosure Block: **K. De Andrade:** None.

Li-Fraumeni syndrome (LFS) is a highly penetrant cancer predisposition syndrome caused by germline pathogenic *TP53* variants associated with very high risks of cancer starting in early childhood throughout the entire lifespan. There are limited data on the risk of cancer in LFS compared with the general population, on genotype-phenotype correlations, or on potential genetic modifiers of *TP53* function. This study included 480 carriers of pathogenic/likely pathogenic (P/LP) germline *TP53* variants participating in an IRB-approved study (NCT01443468). Median age at last follow-up was 36.7 years for females and 34.3 years for males. Variants were categorized based on loss-of-function (LOF) and dominant-negative effect (DNE) properties. LFS-associated cancer incidences were compared with the general population using the Surveillance, Epidemiology, and End Results (SEER) 1975-2017 registry. Cancer incidences were evaluated using family-clustered Cox-regression models. Eight polymorphisms in *TP53, MDM4,* and *MDM2* were evaluated as possible first-cancer incidence modifiers. Three hundred five (63.5%) participants were diagnosed with at least one cancer. The standardized incidence of any cancer in individuals with LFS compared with age and sex-matched SEER data peaks from childhood to age 30 (> 60-fold higher), remaining nearly 10-fold higher after age 50. Breast cancer accounted for 56.9% of all female first-primary cancers with a lifetime cumulative first-cancer incidence of approximately 56%. Median age at first cancer in women substantially increased by censoring breast cancer cases (59.4 *versus* 33.7 years). Regardless of DNE status, variants with LOF properties were associated with significantly earlier ages at first-cancer diagnoses. Individuals whose first cancers were diagnosed before 17 years of age develop a second cancer after a longer time interval than those with first cancers diagnosed at later ages. Multiple cancers were diagnosed within a short time frame in some individuals with LFS suggesting a â□□trigger effectâ□□ phenomenon. The polymorphism rs17878362-A2 showed the highest first-cancer risk (hazard ratio=3.06, pTP53 variant functional groups and common polymorphisms appear to modify median ages at first-cancer diagnoses in LFS. Risk-reducing mastectomy in females with LFS could improve their cancer incidence to parallel that of males with LFS. This study provides important data to develop cancer risk stratification strategies for individuals with LFS.

PrgmNr 2757 - *USP8* gene may contribute to the development of bilateral adrenal hyperplasia and ACTH-independent Cushing syndrome

View session detail

Author Block: **N. SETTAS**[1], E. Kassi[2], A. Angelousi[2], F. R. Faucz[3], C. A. Stratakis[4]; [1]NIH (NIH), BETHESDA, MD, [2]Natl. and Kapodistrian Univ. of Athens, Athens, Greece, [3]NIH / NICHD, Rockville, MD, [4]NICHD, NIH, Bethesda, MD

Disclosure Block: **N. Settas:** None.

Bilateral adrenocortical hyperplasias (BAHs), including primary pigmented nodular adrenocortical disease (PPNAD), isolated micronodular adrenocortical disease (iMAD) and primary macronodular adrenocortical hyperplasia (PMAH), are rare causes of ACTH-independent Cushing syndrome (CS). PPNAD and iMAD usually present in children or adolescents as multiple small (USP8 gene variants. *USP8* is mainly known for being mutated in Cushing disease but as a deubiquitinase it may be involved into other pathways such as Wnt/b-catenin and/or PKA.The first patient was diagnosed with BAH on prenatal ultrasound and subsequently required bilateral adrenalectomy for CS as she had virilization, hirsutism, hypertension and cardiac hypertrophy 9 weeks old. Adrenalectomy revealed that she had iMAD. She also presented with hemihypertrophy of the right leg, labia and mild newborn hypoglycemia, however she was negative for Beckwith-Wiedemann mutation. Gene analysis of *PRKAR1A* did not reveal any mutations. After whole exome sequencing (WES), we found a novel heterozygous *USP8* variant (c.1387_1393delinsT, p.A463_I465delinsF) at germline level and loss of heterozygosity (LOH) at tumor level. Immunohistochemistry showed significantly lower expression of USP8 protein in both of her adrenals compared to a control tissue. Immunofluorescence showed higher expression of PRKACA and PRKACB proteins compared to control tissue. The second case is a 59-year old female with osteoporosis who failed to suppress cortisol levels after low dose dexamethasone administration. MRI revealed an adenoma on the right adrenal (2.6cm). She underwent right adrenalectomy and was found to have PMAH. We performed WES in germline level and we detected a novel heterozygous missense *USP8* variant (c.287A>G, p.K96R) that is present also at tumor level. Immunohistochemistry showed significantly lower expression of USP8 protein in her adrenal tumor compared to the control tissue. Immunofluorescence showed higher expression of PRKACA and PRKACB proteins compared to control tissue. No LOH was identified. These preliminary findings support *UPS8* involvement in the development of adrenocortical disease. We are currently performing further in vitro studies to evaluate the effect of these two *USP8* variants into the canonical Wnt pathway and PKA pathway which are commonly involved in adrenocortical disorders.

PrgmNr 2758 - Cancer Risk B is a rapid and accurate test of peripheral blood mononuclear cells for prediction of breast cancer risk

View session detail

Author Block: **J. Loke**[1], H. Ostrer[1], S. D. Klugman[2], D. Gruber[3], D. H. Tegay[4], I. Alim[5], K. Onel[6], A. La Bella[7], S. Yam[8], L. C. Xia[1]; [1]Albert Einstein Coll. of Med., Bronx, NY, [2]Montefiore Med. Ctr., Bronx, NY, [3]Columbia Univ, New York, NY, [4]Northwell Hlth., Mount Sinai, NY, [5]Morgan And Mendel Genomics, Bronx, NY, [6]Univ Chicago, Chicago, IL, [7]Northwell Hlth., Lake Success, NY, [8]Morgan and Mendel Genomics, Inc., Bronx, NY

Disclosure Block: **J. Loke:** Major Stockholder/Ownership Interest; Morgan and Mendel Genomics, Inc..

Identifying women at high risk for developing breast cancer can trigger a personal program of surveillance for early detection, prophylactic surgery, or chemoprevention to reduce risk. Cancer gene panel sequencing yields risk-modifying results in only 15-20% of cases, suggesting the need for a new approach. Cancer Risk B (CR-B) is a rapid test (2 days) that uses flow variant analysis to assess the effects of variants in the DNA double strand break (DSB) repair pathway on the phosphorylation of p53 and nuclear localization of BRCA1 and BRCA2. CR-B was compared in matched peripheral blood mononuclear cells (PBMCs) and lymphoblastoid cells (LCLs; N=20), then applied to two groups of subjects who received coincidental gene panel sequencing to assess sensitivity, specificity and accuracy, and utility for annotating variants of uncertain significance (N=159). When comparing matched LCLs and PBMCs and inter-day tests, CR-B yielded highly reproducible results (r^2>0.9). The CR-B phenotype demonstrated a bimodal distribution: CR-B+, indicative of DSB repair defects, and CR-B-, indicative of wild-type repair**.** When applied to subjects with known pathogenic variants (positive controls) or negative relatives of known pathogenic variant carriers (negative controls), the sensitivity of the test was 91% and the specificity was 96%. The CR-B+ phenotype was rescued by gene transfer using wild-type cDNA expression plasmid transduction into LCLs or CRISPR-edited cells. As a functional genomics test with strong evidence, the CR-B- phenotype predicted VUS as benign or likely benign. CR-B could represent a rapid and accurate alternative to panel sequencing for identifying women at high-risk for cancer and is a useful functional genomics test for predicting VUS as benign.

PrgmNr 2759 - Casting a Wide Net: Finding actionable results in non-breast cancer genes on multi-gene panel testing in a breast cancer cohort

View session detail

Author Block: **M. Rohanizadegan**[1], L. Kipnis[1], S. Stokes[1], B. L. Bychkovsky[1], R. G. Scheib[1], H. Q. Rana[2], J. E. Garber[3]; [1]Dana Farber Cancer Inst., Boston, MA, [2]Dana-Farber Cancer Inst., Boston, MA, [3]Dana Farber Cancer Inst, Boston, MA

Disclosure Block: **M. Rohanizadegan:** None.

Background: Multi-gene panel testing (MGPT) for hereditary cancer syndromes allows for concurrent analysis of genes associated with various cancer types. This may lead to the identification of unexpected pathogenic/likely pathogenic variants (P/LPVs) in high penetrance cancer genes with no link to the original cancer. In this study we examined the landscape of P/LPVs in a breast cancer (BC) cohort who underwent MGPT and assessed clinical and family history features consistent with identified P/LPVs.

Methods: We retrospective reviewed patients seen at a single institution who underwent MGPT from 1/1/15- 5/24/21. MGPT was defined as additional gene testing beyond the 9 genes associated with BC (*ATM, BRCA1, BRCA2, CDH1, CHEK2, PALB2, PTEN, STK11, TP53*). Subjects with P/LPVs in moderate-penetrance BC genes for which cancer risk is not established yet, heterozygous P/LPVs in genes with a recessive inheritance pattern and those with *APC* I307K and *FH* p.Lys477dup were excluded. Deidentified pedigrees were analyzed to determine whether there was clinical suspicion of a P/LPV in any of the non-BC high-penetrance genes considering national testing guidelines or clinical diagnostic criteria. **Results**: Among 7086 patients, 888 (12%) were found to have at least one P/LPV in a cancer susceptibility gene. We identified 159 P/LPVs in genes not typically associated with risk for BC in 157 (2%) patients; among these, there was clinical suspicion for the identified P/LPV in only 51 patients (32%, including 45 with Lynch syndrome, 33 with ovarian cancer related genes i.e *BRIP1, RAD51C* and *RAD51D*, 18 with SDHx, and 17 with melanoma related genes such as *CDKN2A, BAP1* and *MITF*). Medical or surgical management would be affected by the MGPT result in 93% of subjects. Only 42% of subjects with Lynch syndrome had a clinical suspicion based on personal or family history, 11% with SDHx related hereditary cancer syndrome and 18% of subjects with genetic risk for ovarian cancer. **Conclusion**: Of 7086 BC patients who underwent MGPT, 2% were found to have a P/LPV in a high-penetrance non-BC gene that would have been missed by a smaller BC gene panel. In all cases, cascade testing was offered to at-risk family members, allowing for cancer and reproductive risk stratification and management. This study estimates the prevalence of P/LPVs from MGPT in a referral center, which may address some concerns about the chance of unanticipated findings with comprehensive testing in the breast cancer population.

PrgmNr 2760 - Cellular diversity of human breast tissue evaluated in Pre-templated Instant Partitions for single-cell RNA sequencing (PIPseq)

View session detail

Author Block: **J. Zhang**[1], K. Phong[2], Z. Gartner[2], R. Meltzer[1], K. Fontanez[1]; [1]Fluent BioSci., Watertown, MA, [2]Univ. of California, San Francisco, San Francisco, CA
Disclosure Block: **J. Zhang:** Salary/Employment; Fluent BioSciences.
Single-cell RNA sequencing (scRNA-seq) has made profound impacts in the study of cellular and molecular diversity in complex tissues. Breast tissue comprises a diverse mixture of epithelial, lymphatic, vascular, and immune cell populations, and the structure and composition of breast tissue remodels continuously throughout a womanâ□□s lifetime. Previous scRNA-seq analysis of banked reduction mammoplasty tissue has revealed changes in cell type abundances in response to physiological events such as childbirth and the menstrual cycle. Performing this analysis on a larger patient cohort will strengthen these correlations and identify further links between physiological factors and breast tissue cellular composition. However, current scRNA-seq platforms make it difficult and cost-prohibitive to process a large cohort of asynchronously obtained patient samples. Fluent BioSciences has developed Pre-templated Instant Partitions (PIPs) to simultaneously segregate complex cell mixtures into partitions with barcoded template particles that can be easily processed for scRNA-seq (PIPseq) without the need for complex instrumentation or microfluidic consumables, allowing PIPseq to be performed in any moderately equipped laboratory at a lower cost per cell than other commercially available scRNA-seq platforms. As an initial validation of PIPseq for the study of breast tissue, approximately 1000 primary human mammary epithelial cells were processed by PIPseq. Clustering analysis indicates 2 primary cell clusters, and marker gene analysis of Keratin 14 and 19 clearly segregates these populations to myoepithelial and luminal cell populations. This is in agreement with cell populations previously characterized with other scRNA-seq methods. PIPseq has the required biological sensitivity to identify cell populations in primary breast tissue samples. Future efforts will correlate cell populations in primary resected breast tissue with patient-matched estrogen and progesterone levels.

PrgmNr 2761 - Improving predictions of variant functional effects using deep mutational scanning data

View session detail

Author Block: **T. Yu**[1], J. Fife[1], R. Sherwood[2], C. Cassa[3]; [1]Brigham and Women's Hosp., Cambridge, MA, [2]Brigham and Women's Hosp., Harvard Med Sch., Boston, MA, [3]Harvard Med. Sch., CAMBRIDGE, MA

Disclosure Block: **T. Yu:** None.

Predicting the functional impact of coding variants remains a substantial challenge, with applications in population genetics, molecular biology, and clinical genetics. Computational predictive tools often rely on features related to structure, conservation, or population data. To complement these predictions, deep mutational scanning (DMS) methods are used to measure the functional effects of allelic variants. These *in vitro* screens are collectively informative, but may be individually noisy per variant.

We aggregate DMS data to identify regions of proteins likely to be functionally impactful versus regions more tolerant to mutation. We do this by pooling variant functional effect scores in a moving window. Importantly, we restrict our moving averages to a) amino acid substitutions that could possibly occur, and b) to the 80 amino acid substitutions that are most likely to be informative, by excluding substitutions that are often damaging or rarely damaging.

We then train a 2-state Hidden Markov Model (HMM) using this moving average, the phyloP 100-way score, and a positional score which represents the predicted functional effect from DMS data. The predicted states of the HMM then provide a regional â□□scaffoldâ□□ of functional impact for a mutation in the region. Importantly, no clinical diagnostic data was used to develop this scaffold.

Finally, we make a context-sensitive functional score for any SNV based on its HMM assignment and the specific amino-acid substitution. We generate the amino acid substitution matrix using DMS data from 11 genes other than *BRCA1*. Using lab-classified pathogenic and benign mutations in *BRCA1* from ClinVar, we can significantly separate these variants as pathogenic or benign based on the contextual score (p-6). In contrast, the normalized BLOSUM62 substitution matrix can only achieve pThis project provides two important contributions: 1) we are able to aggregate signal from DMS data to identify regions of higher or lower functional impact, and 2) by restricting to 80 amino acid substitutions, we are able to provide highly significant functional scores with substantially fewer DMS assays. Future work will integrate these functional scores with existing predictive features.

PrgmNr 2762 - Large-scale functional assessment of *RAD51C* missense variants with PARP inhibitor screens

View session detail

Author Block: **G. Montalban**[1], L. Milano[1], A. Rodrigue[1], Y. Coulombe[1], S. Desjardins[1], M. Dumont[1], C. Joly-Beauparlant[1], P. Soucy[1], K. Matreyek[2], J-Y. Masson[1], J. Simard[1]; [1]CHU de QuÃ©bec-UniversitÃ© Laval Res. Ctr., Quebec city, QC, Canada, [2]Case Western Reserve Univ. Sch. of Med., Cleveland, OH
Disclosure Block: **G. Montalban:** None.
Germline pathogenic variants in homologous recombination repair (HRR) genes are associated with increased risks of developing breast/ovarian cancer. RAD51C participates in HRR by interacting with other RAD51 paralogues to promote the stabilization of the RAD51 nucleoprotein filament in the homologous DNA strand. Large case-control studies have confirmed that *RAD51C* loss-of-function variants confer moderate risks of developing breast/ovarian cancer (Dorling L et al., NEJM 2021; Hu C et al., NEJM 2021). However, the increased identification of variants of unknown clinical significance (VUS) hampers genetic counseling of breast/ovarian patients and families. Here we present a massively parallel functional approach to study the molecular impact of all missense substitutions in the *RAD51C* gene using PARP inhibitor sensitivity as a readout. A *RAD51C* mutagenesis library was designed to cover all possible missense substitutions (~7,500 variants; 98% coverage). The library was barcoded and cloned into an inducible, recombinase-site containing vector. Landing pad cell lines were generated to allow the systematic integration of the library and to ensure the recombination of one variant per cell, keeping the genotype-phenotype link. Briefly, endogenous RAD51C expression was silenced using siRNAs targeting the 3â□□UTR region of the gene and cells were exposed to different PARP inhibitors. Genomic DNA from untreated and treated pools was extracted and sequenced in a NovaSeq instrument. To date, several â□□mini-screensâ□□ have been performed covering different regions of the gene. All variants were detected in the pre-treated pools at a similar abundance, confirming their optimal integration and expression. Variant read counts were reduced for known deleterious variants (positive controls) after treatment, confirming the synthetic lethal effect of PARPi when RAD51C is not functional. Whole screens with the barcoded library are ongoing, as well as the application of specific computational tools for the calculation and calibration of variant fitness scores (Enrich2 and DimSum; Rubin et al., 2027; Faure AJ et al., 2020, respectively). We have developed a large-scale functional approach to measure the impact of all missense variants in the *RAD51C* by screening for PARPi sensitivity. Future work will focus on validating our functional scores with published works, clinical databases and complementary functional assays. The final goal is to improve the interpretation of *RAD51C* VUS and accelerate their clinical translation.

PrgmNr 2763 - Measuring (almost) all possible*STK11/LKB1* missense variants in monogenic disease and cancer

View session detail

Author Block: **D. Zimmerman**, A. Cote, N. Kishore, M. Gebbia, W. Loggerenberg, J. Weile, F. Roth; Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **D. Zimmerman:** None.
Measuring the Functional Impact of (Almost) All Possible *STK11/LKB1* Missense Variants in Monogenic Disease and Cancer Germline missense variants in Serine/Threonine Kinase 11 (*STK11*, aka *LKB1*) cause Peutz-Jeghers Syndrome (PJS), a rare disease that increases lifetime cancer risk. Somatic *STK11* variants also contribute to multiple cancers (including ~30% of lung cancers). Early genetic diagnosis of PJS helps establish life-long disease management, in part by identifying patients for pre-emptive surgeries. *STK11*-deficient lung cancers are resistant to widely used immune checkpoint inhibitors and proactive detection would limit prescription of an ineffective therapy. Unfortunately, nearly 90% of clinically interpreted *STK11* missense variants are classified as Variants of Uncertain Significance (VUS). Resources for variant interpretation are sorely needed, the American College of Medical Genetics has listed *STK11* as one of their 59 clinically actionable genes. Since functional assays provide strong evidence for clinical variant interpretation, we aim to test all possible *STK11* missense variants. To this end, we have employed two scalable assays for *STK11* variant function: 1) complementation of a S. cerevisiae strain lacking the yeast orthologues *SAK1*, *TOS3*, and *ELM1*; and 2) toxicity of *STK11* expression in the human HeLa cell line. We describe assay validation, mutagenized library construction, and other progress towards a *STK11* missense variant effect map, potentially improving diagnosis and therapy for PJS and *STK11*-related cancers.

PrgmNr 2764 - Pathological variant c.8143C&gtT in *ATM* gene in a patient with non-Hodgkin lymphoma

View session detail

Author Block: **I. Cuero-Quezada**[1], L. Bobadilla-Morales[1], M. Orozco-Vela[1], S. A. BRUKMAN[2], A. Corona-Rivera[1]; [1]Univ. of Guadalajarada, Guadalajara, Mexico, [2]UNIVERSIDAD DE GUADALAJARA, GUADALAJARA, Mexico
Disclosure Block: **I. Cuero-Quezada:** None.
Introduction. Lymphomas belong to the group of lymphoid neoplasms, the prognosis depends on the histological type, clinical factors and molecular characteristics. Non-Hodgkin's lymphoma is a malignant neoplasm of B, T or NK cells, it can infiltrate lymphoid and hematopoietic tissues, the main symptoms are night sweats, fever and weight loss, the diagnosis is established through a tissue biopsy. Hodgkin's lymphoma is a B-cell lymph node neoplasm, the neoplastic cells are known as Reed-Sternberg cells. **Objective.** Present a case with an initial diagnosis of non-Hodgkin's lymphoma and relapsed Hodgkin's lymphoma presenting the c.8143C>T germinal variant in the ATM gene. **Clinical report.** 18-year-old female, normodevelopmental pregnancy, without alterations and adequate prenatal control, normal psychomotor development. Family history of CNS tumors. It begins with symptoms at 14 years of age with upper respiratory symptoms presenting fever of 39Â°C and cough, presence of pain in the left jaw, tooth extraction is performed, it is diagnosed as a histiocytoma with partial improvement, at 15 years of age it presents adynamia, weight loss, larger cervical nodes, a biopsy is taken obtaining the diagnosis of non-Hodgkin lymphoma of T cells with markers CD20(-), CD3(+), CD30(+), CD15(+), PML-1(-), Langerginas(-), MUM-1(-). The evolution was stable and remained in maintenance until the age of 17 when she presented dyspnea, B symptoms and generalized paleness, a relapse was confirmed by CT of the chest and abdomen, a mediastinal mass biopsy was taken with a result of CD20 mixed cellularity Hodgkin lymphoma (+) in 80% of Reed-Sternberg cells.Material and method. Conventional karyotype of the bone marrow had a normal result, with a family history of cancer and the presence of two primary tumors a search for mutations in familial cancer genes was performed using the CentoCancerÂ® panel. **Results.** Through sequencing, the variant c.8143C>T in the ATM gene (Leu2715Phe) was detected, classified as a variant of uncertain significance.Discussion. Mutations in ATM have been related to the risk of developing cancer, especially lymphoid neoplasms, it is suggested that ATM act as a tumor suppressor. The variant is classified as of uncertain significance, however, due to the antecedents, we consider it to be a pathological variant. Pathogenic variants have been reported in tumor cells and germ cells in various types of leukemia and lymphoma. The search for variants in ATM are important to help the risk stratification of patients who present them. **Conclusions.** This work makes it possible to add the ATM gene variant c.8143C>T as a pathogenic classification.

PrgmNr 2765 - Uncovering the functional variants and target genes of the 7q32 pancreatic cancer risk locus

View session detail

Author Block: **A. O'Brien**[1,2], J. Hoskins[1], D. Eiser[1], K. E. Connelly[1], J. Zhong[1], I. Collins[1], Pancreatic Cancer Case-Control Consortium, Pancreatic Cancer Cohort Consortium, S. J. Chanock[1], A. P. Klein[3], H. E. Arda[1], L. T. Amundadottir[1]; [1]NIH, Bethesda, MD, [2]Queen's Univ. Belfast, Belfast, United Kingdom, [3]Johns Hopkins Univ. Sch. of Med., Baltimore, MD
Disclosure Block: **A. O'Brien:** None.
Pancreatic ductal adenocarcinoma (PDAC) is set to become second leading cause of cancer mortality over the next decade. PDAC has a 5-year survival rate of only ~9%, necessitating a better understanding of its etiology. PDAC risk factors are both environmental and genetic, with heritability estimated at around 20%. However, highly penetrant coding variants are rare in the population and account for as little as 1% of the phenotypic variance for PDAC.

To identify common PDAC susceptibility alleles, genome-wide association studies (GWAS) in people of European ancestry have identified 17 loci at which germline variants associate with PDAC risk. Mostly non-coding, many of these variants are expected to confer risk through allele-specific changes in transcription factor binding, leading to altered target gene expression over a lifetime. The identification of functional variants from GWAS is complicated by the correlation of alleles through linkage disequilibrium (LD).

Here, we leveraged Bayesian fine mapping to identify a credible set (CS) of seven single nucleotide polymorphisms (SNPs) in high LD at the 7q32.3 risk locus. CS variants were annotated with epigenomic datasets we previously generated from a range of pancreatic cell lines, organoids and primary patient samples, encompassing histone modification marks (ChIP-seq), open chromatin (ATAC-seq) and chromatin conformation capture (Capture-HiC, HiChIP). Two variants from the CS (rs6971499 and rs6970779) were found to consistently lie within regions annotated as super-enhancers and marked by open chromatin. Electrophoretic mobility shift assays (EMSA) illustrate allele-specific protein binding for rs6971499-T and rs6970779-G in PANC-1 and MIA PaCa-2 cell lines, while luciferase reporter assays in the same cell lines demonstrate that both alleles drive significant changes in transcriptional activity.

Both SNPs exhibit chromatin looping to the promoters of *LINC-PINT* and *MIR29a*, representing two plausible target genes. *MIR29* family members are implicated in both suppression and progression of many cancers and are key regulators of PDACâ□□s characteristic stromal presentation. *LINC-PINT*, a p53-induced long non-coding RNA is associated with epigenetic silencing via recruitment of the PRC2 polycomb repressive complex.

Ongoing proteomic experiments aim to identify the transcription factors differentially bound to these alleles, while CRISPR is being used to validate enhancer variants in their native chromatin context. Future experiments may further characterize these transcription factors and their target genes with an eye toward better understanding their roles in PDAC carcinogenesis.

PrgmNr 2766 - 3D confirmation analysis of the *FGFR2* kinase domain with disease-associated mutations reveals novel features that correlate with clinical presentation

View session detail

Author Block: **Y. Lian**[1], D. Bodian[2], A. Shehu[1]; [1]George Mason Univ., Fairfax, VA, [2]Inova Translational Med. Inst., Falls Church, VA
Disclosure Block: **Y. Lian:** None.
Receptor tyrosine kinases (RTKs) play an essential role in the regulation of many cell signaling processes including cell proliferation, angiogenesis, and survival. Most members of the RTK family are transmembrane proteins that assume different conformations to perform their cellular functions. These conformational dynamics are physiologically regulated by binding ligands to the extracellular domain and post-translational modifications such as receptor activation by autophosphorylation of the intracellular kinase domain and are also affected by pathogenic mutations. Activation is tightly regulated, with dysregulation by germline or somatic mutation leading to activation in the absence of ligand binding and causing various cancers and developmental disorders.
Fibroblast growth factor receptor 2 (*FGFR2*) kinase has been a model for studying the mechanism by which gain of function mutations in the tyrosine kinase domain cause activation in the absence of ligand binding. Germline mutations in the kinase domain of *FGFR2* lead to skeletal disorders, primarily dwarfism, and craniofacial malformation syndromes. *FGFR2* has also been implicated in cancers including breast and gastric cancer. Many X-ray and NMR structures of inactive and activated forms, and of isoforms with activating or inhibiting mutations, are now available. Analysis of these structures has led to the proposal of a molecular brake preventing the ATP-binding A-loop from adopting the activated conformation. Activating mutations can relieve the brake, affecting a long-range allosteric process that causes a change in equilibrium between the active and inactive states. However, this fundamentally dynamic process is not well captured by the static structures and so remains not well characterized.
In this study, we employ all the available structures of *FGFR2* kinase from the Protein Data Bank (PDB) to capture and characterize the conformational dynamics of the *FGFR2* kinase domain and correlate these dynamics with known functional regions and disease type. We use machine learning methodologies to elucidate the relationship between conformational changes and the kinaseâ□□s activation region that occurs upon phosphorylation. We show that linear and nonlinear representation learning techniques agree on the characterized dynamics. Organizing structures by the revealed intrinsic motions and activated region and disease type shows good co-localization, linking mutations to dynamics and clinical presentation. Our work demonstrates the value of including machine-learned representations of conformational dynamics in identifying RTK regulatory functions and associated disease types.

PrgmNr 2767 - A genome-first approach characterizing the prevalence and cancer phenotype of pathogenic germline *DICER1* variants in two large, unselected cohorts

View session detail

Author Block: **J. Kim**[1], J. Haley[2], H. Rao[2], U. T. Mirshahi[2], Geisinger-Regeneron DiscovEHR Collaboration, K. Schultz[3], D. J. Carey[2], D. Stewart[4]; [1]Natl. Cancer Inst., Rockville, MD, [2]Geisinger Clinic, Danville, PA, [3]Intl. Pleuropulmonary Blastoma/DICER1 Registry, Minneapolis, MN, [4]Natl. Cancer Inst, Rockville, MD

Disclosure Block: **J. Kim:** None.

DICER1 is an RNaseIII endonuclease crucial for processing pre-microRNA into mature, active microRNA (miRNA). Loss-of-function germline variants in *DICER1* underlie an autosomal dominant disorder with an increased risk of numerous neoplasms, macrocephaly, and thyroid-related disease. The availability of large-scale, exome-sequenced cohorts with linked electronic health records (EHR) from the DiscovEHR study and UK Biobank cohort permits genome-first exploration of phenotype in individuals with a pathogenic germline *DICER1* variant. We previously observed *DICER1* pathogenic germline variants in 1 out of 4600 exomes and an increased risk of thyroidectomy and thyroid cancer in the DiscovEHR study of ~92K individuals. In this study, we described the prevalence of pathogenic germline *DICER1* variants in exomes of 175K individuals from the DiscovEHR study and 200K individuals from the UK Biobank cohort.*DICER1* variants were classified as pathogenic if it was a RNaseIIIb mutation hotspot, nonsense, frameshift, or a cryptic splice-site, and not located in last exon. For the UK Biobank cohort, we queried for cancer diagnoses and procedures using self-reported questionnaire, cancer and death registry, and EHR. For the DiscovEHR study, we queried the Geisinger Cancer Registry and EHR for similar diagnoses and procedures. We identified 21 individuals (1:8355) and 25 individuals (1:9234) with a *DICER1* pathogenic variant in the DiscovEHR study and the UK Biobank cohort, respectively. In the DiscovEHR study, we identified five people with malignancies, including two lung cancers (54 yrs, 62 yrs), meningioma (41 yrs), thyroid cancer (37 yrs), and breast cancer (57 yrs). Furthermore, of the 21 people, 12 had an ICD10 code for thyroid-related diseases (hypothyroidism, thyrotoxicosis, and/or nontoxic goiter). Age of first diagnosis for thyroid disease ranged from 20-71 yrs (female mean: 39 yrs, male mean: 52 yrs). In UK Biobank, we found three people with malignancies, including pineoblastoma (15 yrs) and two breast cancers (62 yrs and 50 yrs). The two people with breast cancer did not harbor exonic variants in known breast cancer susceptibility genes. In addition, we observed six cases of thyroid diseases (two thyroidectomy, two hypothyroidisms, thyrotoxicosis, non-toxic goiter), ranging from 14-68 yrs (female mean: 51 yrs, male: 14 yrs). In the UK Biobank, there was no ICD code (other than those for thyroid disease) excess in subjects with a *DICER1* pathogenic variant. Here we provide updated prevalence and thyroid penetrance for pathogenic *DICER1* variants in two large cohorts. The observation of breast cancer in the *DICER1* heterozygotes merits additional investigation.

PrgmNr 2768 - An analysis of automated cell-type labelling algorithms for tumour micro-environment single-cell RNA sequencing data

View session detail

Author Block: **E. Christensen**[1,2], P. Luo[3], A. Turinsky[4], M. Husic[4], A. ï»¿Mahalanabis[4], A. Naidas[1,2], J. Diaz-Mejia[3], T. Pugh[3,4,5], A. K. Ramani[4], P. Shooshtari[6,1,7]; [1]Univ. of Western Ontario, London, ON, Canada, [2]Children's Hlth.Res. Inst., London, ON, Canada, [3]Princess Margaret Cancer Ctr., Toronto, ON, Canada, [4]Ctr. for Computational Med., The Hosp. for Sick Children, Toronto, ON, Canada, [5]Dept. of Med. Biophysics, Univ. of Toronto, Toronto, ON, Canada, [6]Children Hlth.Res. Inst., London, ON, Canada, [7]Ontario Inst. for Cancer Res., Toronto, ON, Canada
Disclosure Block: **E. Christensen:** None.
Single cell RNA sequencing (scRNA-seq) allows researchers to precisely define complex cell populations through the examination of gene expression in individual cells from a sample. This is particularly useful in cancer research, as the tumour microenvironment (TME) can be incredibly heterogeneous, containing malignant cells, cells of the tissue of origin, immune cells, fibroblasts, and more. This heterogeneity often affects how cancer progresses and responds to treatment, making it important to accurately identify all cells within the TME. Most current scRNA-seq pipelines involve clustering the cells by differentially expressed genes and assigning cell type labels to each cluster based on gene sets. This labelling step is typically a manual task, which can rely on subjective decisions and may yield inconsistent results. Several automated methods for identifying cell types exist, and have been tested on healthy scRNA-seq samples, but their performance on more complex TME data is unknown.

Here we evaluate 26 methods to automate the process of assigning labels to TME cells - 19 supervised cell-based methods and 7 semi-supervised cluster-based methods. Our study shows that cell-based algorithms outperform cluster-based algorithms overall. This could be because the cell-based algorithms are fully supervised, learning from each individual cell in the training set instead of only examining select genes within a group of cells. Overall scPred, CaSTLe, scVI, and SVM are the top-performing algorithms. Of these algorithms, SVM is the fastest running. Among cluster-based methods, GSVA ranked highest. We also tested the ability of cell-based algorithms to train a classifier on one group of patients and predict the cells in another group and found that algorithms performing well with random training and testing sets also perform well on separate patients, suggesting they would be applicable in a clinical setting.

Cluster-based methods are often unable to identify malignant cells within a dataset due to lack of appropriate gene signatures. Cell-based algorithms, however, tended to perform better on the malignant cell categories - an interesting result as the algorithms learn from the data they are provided and most methods are not tailored to TME data specifically. This may be due to a higher proportion of malignant cells present in the data providing a better source for the algorithms to learn a malignant cell profile. In summary, our analysis presents a set of guidelines for researchers to understand the various cell type identification methods and select the best one for their work, allowing them to create new and improved treatments which ultimately improve patient care.

PrgmNr 2769 - Evaluating Immune Repertoire Patterns During Prostate Cancer Tumor Progression

View session detail

Author Block: **H. Arora**[1], A. Vedenko[1], R. Reddy[1], D. Van Booven[2]; [1]Dept of Urology, Univ. of Miami, Miami, FL, [2]Univ Miami Miller Sch Med, Miami, FL
Disclosure Block: **H. Arora:** None.
Introduction: Prostate cancer (PCa) progression from benign androgen (AR) dependent stage to AR independent castration resistant and then a neuroendocrine stage is often accompanied by resistance to therapy.. This resistance calls for ways that can help in understanding the etiology of progression in order to improve outcome.. Similarly, Immune checkpoint inhibition (ICI) has revolutionized treatment of many tumor types, including prostate cancer. Such therapies focus on restoring antibody diversity to eliminate tumor cells that evaded immune detection, but is effective only in a subset of patients.. This limited efficacy be in part from tumor heterogeneity and the unique ability of PCa to evolve. Therefore, this study focuses on evaluating the immunomodulatory patterns that are correlated to stages of PCa progression. **Methods**: The analysis contains 3 parts: PCa subclassification, mutation analysis, and Ig expression. First, we obtained all RNAseq from the PRAD project contained in the TCGA (n=500). A general PCA was performed, and then patients were stratified into subclasses by using gene signatures obtained from the Disease/Gene Database (disgene.net). The results were then classified using basic clustering to refine these genes, and finalize the subclasses. Second, mutations were downloaded for the same PRAD TCGA dataset. Mutation types as well as number of mutations were classified per patient in order to come up with a mutational profile per patient. These profiles were then used to scan for mutations across severity by Gleason score (GS). Thirdly, Ig profiles were evaluated using the software packages TRUST and MiXCR. Total CDR3 sequences as well as AA length was analyzed and compared with mutation rates as well as switching events. Finally, all stratifications were compared to known tumor severity by GS. **Results**: We observed patterns of Immunoglobulin (Ig) expression that are co-relatable to a) PCa subclasses such as Prostate Adenocarcinoma, Recurrent PCa, and Invasive PCa, b) mutation landscape (with focus on mutation types SNPs, SVs, and fusions), and c) Ig expression. Also, there are immunomodulatory patterns in IGHG1 and IGHG2 that are correlated with tumor severity and stages of PCa progression. Also, diversity variations exist in Shannon entropy with lower diversity in low grade tumor severity suggesting a significant change in the immune landscape. **Conclusions**: The results suggest there are immune modulatory patterns are correlated with stages of PCa, and their correlation with components such as mutations, GS, subclasses of PCa can help in better understanding PCa progression and design therapeutic strategy more effectively.

PrgmNr 2770 - Identification of the transcriptome signature of Beckwith-Wiedemann Syndrome hepatoblastoma predisposition

View session detail

Author Block: **N. Sobel Naveh**[1], E. M. Traxler[1], K. A. Duffy[1], J. Kalish[1,2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Perelman Sch. of Med. at the Univ. of Pennsylvania, Philadelphia,, PA
Disclosure Block: **N. Sobel Naveh:** None.
Beckwith-Wiedemann Syndrome (BWS, OMIM 130650) is one of the most common epigenetic cancer predisposition syndromes. BWS is caused by alterations to human chromosome 11p15, which contains two independent domains subject to genomic imprinting. Imprinting Center 1 (IC1) encompasses *Insulin-like Growth Factor 2 (IGF2)* and the non-coding growth suppressor *H19*; Imprinting Center 2 (IC2) encompasses the *potassium voltage-gated channel subfamily Q member 1 (KCNQ1),* its antisense transcript *KCNQ1OT1*, and cell-cycle regulator *Cyclin Dependent Kinase 1C (CDKN1C)*. Patients with BWS caused by isolated loss of methylation (LOM) at IC2 or loss-of-heterozygosity due to paternal uniparental isodisomy (pUPD11) encompassing both domains are more likely to develop hepatoblastoma (HB) and patients with BWS have at an estimated 2,200 times greater risk of HB tumorigenesis. While 11p15 epigenetic changes are components of this pre-cancer genome, their downstream molecular effect and the mechanism by which they cause oncogenesis has not been previously investigated. Using clinical data and patient samples from the International BWS Registry and Biorepository, we performed RNA sequencing (RNA-Seq) on BWS HB tumors from patients with IC2 LOM or pUPD11. We identified 610 differentially expressed genes between the HB and non-tumor samples. We identified an additional 139 differentially expressed genes that comprise a BWS HB predisposition signature; these genes are over- or under-expressed in both BWS tumor and BWS matched normal samples compared to control livers. These genes largely represent members of cell cycle, chromatin organization, and response to insulin regulation, as well as the following signaling pathways: WNT, MAPK, AKT, and HIPPO. To better understand the potential relationship between these genes, we visualized their protein-protein interaction (PPI) network. Distinguishing this profile from non-syndromic tumors, proto-oncogenes such as *HRAS, SHC1,* and *WNT1* are under-expressed. In contrast, over-expressed hub genes *HDAC1, MAPK8, IL2*, and *TGFBR2* may represent future targets for BWS-associated HB therapeutic development. The central hub gene MYC has been observed to be regulated by *IGF2* and *KCNQ1OT1* levels in other cell types; as such we present a model wherein these 11p15 genes modulate *MYC* in BWS liver cells to initiate overgrowth and HB predisposition which then dysregulates the aforementioned pathways.

PrgmNr 2771 - Inferring cancer evolution from single tumour biopsies using synthetic supervised learning

View session detail

Author Block: **T. Ouellette**[1,2], P. Awadalla[1,2]; [1]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **T. Ouellette:** None.
A significant hurdle in understanding cancer evolution *in vivo* are the constraints surrounding serial sequencing, through space or time. For this reason, tumour biopsies are primarily sequenced in bulk from a single site and at a single time point. Although multi-region and single-cell data are becoming increasingly utilized, single time point, bulk sequenced biopsies still represent the major data source for studying cancer genomics and evolution in individual patients. Given this limitation, a reasonable strategy for inferring evolution in single tumour biopsies has been to utilize theoretical population genetics to capture signatures of selection from the variant allele frequency (VAF) distribution. The premise being that fitness-altering mutations will deterministically change in frequency over time, leading to characteristic and quantifiable deviations in the VAF distribution relative to some neutral evolutionary scenario. VAF-based methods have been employed to differentiate between positive selection and neutral evolution, to examine growth patterns, to quantify subclonal fitness and time subclonal emergence, and to build population genetics informed mixture models that account for neutral dynamics, that shape, to some extent, all tumour populations. With that said, existing VAF-based methods, although mechanistic and useful in many cases, have apparent limitations. For example, single statistics are compressive and cannot infer complex information, approximate Bayesian computation methods suffer from the curse of dimensionality and can be prohibitively slow due to a rate-limiting simulation step required for each sample, and mixture models, used to identify subclonal populations, are subject to incorrect clustering in the presence of excess dispersion or sparsity. **Here, we develop a synthetic supervised deep learning approach to infer cancer evolution in bulk sequenced single tumour biopsies using only VAF information.** By generating synthetic VAFs, as a proxy for ground truth, from plausible simulations of tumour evolution, we can train deep learning models to classify and quantify evolutionary parameters in real patient tumours. Overall, a synthetic supervised learning approach allows us to avoid information loss that comes with compressing data into a single statistic prior, to separate simulation and training from prediction which ultimately makes inferences faster, and to learn concise representations of the underlying data leading to more accurate predictions relative to existing methods. We apply our synthetic supervised learning approach to over 150 whole-genome sequenced (WGS) single tumour biopsies.

PrgmNr 2772 - Long-read sequencing reveals alternative splicing of transposable elements in breast cancer

View session detail

Author Block: **A. Nesta**[1,2], D. Veiga[2], O. Anczukow[2,1,3], J. Banchereau[2], C. R. Beck[4,2,3]; [1]Genetics and Genome Sci., Dept. of Genetics and Genome Sci., Univ. of Connecticut Hlth.Ctr., Farmington, CT, [2]The Jackson Lab. for Genomic Med., Farmington, CT, [3]Inst. for Systems Genomics, Univ. of Connecticut Hlth.Ctr., Farmington, CT, [4]Dept. of Genetics and Genome Sci., Univ. of Connecticut Hlth.Ctr., Farmington, CT
Disclosure Block: **A. Nesta:** None.
Transposable Elements (TEs) comprise more than half of our DNA and play an increasingly important role in gene expression. On the transcriptome level, transposons contain promoters, splice donors, splice acceptors, and polyadenylation signals which can alter splicing and expression of nearby genes. Many cancers exhibit genome-wide hypomethylation and widespread transcriptome alterations, which may promote the usage of TEs in novel transcripts that are not typically observed in normal tissues. Traditional short-read RNA-Seq methods have identified alternatively spliced TEs in cancer associated genes, but are unable to capture complete isoforms. Short reads may also miss events due to mapping limitations and the repetitive nature of transposons. To comprehensively examine the effects of TEs on cancer transcriptomes, we sequenced 30 breast tissue samples using long-read RNA-Seq (Iso-Seq) and identified >142,000 quality-controlled transcripts containing >84,000 novel alternatively spliced isoforms. We quantified alternative splicing of our Iso-Seq transcriptome using >1000 breast cancer RNA-Seq samples from The Cancer Genome Atlas (TCGA) and the Genome Tissue Expression Project (GTEx) and identified ~3500 tumor-specific splice events (Veiga and Nesta *et al.* BioRxiv 2020). Alternatively spliced TEs were identified by intersecting our data with RepeatMasker. We found that >70,000 (~49%) of our quality-controlled transcripts overlap with TEs and 275 (192 novel) splice events are more prevalent in human breast tumors than in adjacent normal samples or GTEx. Furthermore, 88 of our 275 splice events reside in breast cancer associated genes and may be functionally important. *Alu* transposons in the 3â□□ UTR of transcripts are subject to adenosine to inosine (A -> I) modifications via an RNA editing mechanism shown to promote breast cancer progression. We found 208 previously unannotated 3â□□ UTRs which exhibit A -> I editing in our Iso-Seq transcriptome. Polymorphic retrotransposition events can also impact gene expression. We queried the unmapped segments of our Iso-Seq reads and identified transposon insertions that may impact expression of breast cancer associated genes. Our computational findings have identified putatively disease relevant transcripts; we are currently validating these findings with in vitro assays to establish novel cancer biomarkers and to investigate the potential pro-oncogenic function of our novel TE-containing isoforms. Our research provides a comprehensive window into how TEs affect cancer transcriptomes, and the myriad ways they can influence gene expression.

PrgmNr 2773 - Multiple germline events lead to cancer development in patients with Li-Fraumeni syndrome

View session detail

Author Block: **V. Subasri**[1,2,3], N. Light[1,2], N. Kanwar[1], J. R. Hansford[4], K. E. Nichols[5], K. C. de Andrade[6], P. P. Khincha[6], S. A. Savage[6], J. Schiffman[7], A. Villani[1], A. Shlien[1], A. Goldenberg[1,2,3,8], D. Malkin[1,2]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Vector Inst., Toronto, ON, Canada, [4]Royal Children's Hosp. Melbourne, Melbourne, Australia, [5]St Jude Children's Res. Hosp., Memphis, TN, [6]Natl. Cancer Inst., Bethesda, MD, [7]Univ. of Utah, Salt Lake City, UT, [8]CIFAR: Child and Brain Dev., Toronto, ON, Canada, Toronto, ON, Canada
Disclosure Block: **V. Subasri:** None.
Li-Fraumeni syndrome (LFS) is an autosomal dominant cancer-predisposition syndrome associated with pathogenic germline variants in the *TP53* tumour suppressor gene, in approximately 70% of cases. The early identification of genetic susceptibility is especially important in these patients, as they are at markedly increased risk of developing a spectrum of early-onset malignancies. The impact of early identification of germline cancer-causing aberrations in these patients is substantial, as it informs the prospective management of patients and their families. As such, it is imperative to identify cancer-causing aberrations in the remaining ~30% of patients that fit the clinical definition of LFS, but lack a pathogenic germline variant in *TP53*. In addition, even among LFS patients that harbour a pathogenic germline *TP53* variant, the cumulative lifetime risk of developing cancer is approximately 68% in males and 93% in females. The incomplete/variable penetrance, suggests the presence of additional genetic and epigenetic driver events that contribute to increased or decreased cancer risk in certain individuals. Candidate gene approaches have been considered to evaluate other risk factors; however, few studies have utilized unbiased genome-wide DNA sequencing and epigenomic assays to explain this variability in cancer occurrence.

In this work, we leveraged family-based whole-genome sequencing and methylation of DNA procured from peripheral blood leukocytes to evaluate the germline genetic and/or epigenetic genomes of a large cohort of LFS patients (n=396) who harbour either pathogenic variant (n=374) or wildtype *TP53* (n=22). In patients lacking a pathogenic germline variant in *TP53*, we pinpointed the role of alternative cancer-causing genetic aberrations in 8/14 patients that developed cancer. Among *TP53* variant carriers, 19/49 individuals who developed cancer harboured an additional pathogenic variant in another cancer gene, while variants in the WNT signalling pathway were associated with decreased cancer incidence. In addition, we leveraged the non-coding genome and methylome to identify inherited epimutations that confer increased cancer risk. Overall, our study highlights the immense benefits of expanding genetic and epigenetic testing of LFS patients beyond *TP53*, as *TP53* status alone does not indicate whether an individual will develop cancer.

PrgmNr 2774 - Optical genome mapping reveals novel structural variants in pediatric brain tumors

View session detail

Author Block: **M. Bornhorst**, S. Bhattacharya, Y. Fu, A. Eze, D. Morinigo, J. Nazarian, E. J. Vilain, H. Barseghyan; Children's Natl. Hosp., Washington, DC
Disclosure Block: **M. Bornhorst:** None.
Background: Brain tumors are a leading cause of cancer-related mortality in children. Research shows that structural variants (SVs) play an important role in tumorigenesis and growth. However, most genetic studies have relied on short-read based exome/genome sequencing (SRS) that have limited ability to detect SVs. Optical genome mapping (OGM) utilizes ultra-long DNA molecules to construct and evaluate much longer genomic fragments, making it effective in identifying large SVs. In this study, we utilized OGM for the detection of novel clinically relevant SVs in pediatric brain tumors. Methods: Ultra-high molecular weight DNA was extracted from a cohort of pediatric brain tumors and matched normal samples (n=50). OGM was performed using nanochannel chip arrays on a Saphyr instrument and SVs (large insertions/deletions, inversions, translocations) were identified. We then used *nanotatoR* to subclassify SVs into functional pathways and determined which genes impact specific cell developmental functions. Tumor-associated, low population frequency SVs were identified in all samples. Results: The most common SV events were deletions and translocations. Nearly all samples had an SV that overlapped with a gene known to have clinical significance. *CDKN2A/B* deletions were most common in high grade tumors. OGM also detected clinically relevant, previously unknown fusions (i.e. *NTRK2-KANK1*, *C11orf95-NCOA1* fusion) and deletions (*NF1, ATRX*) that may have affected therapeutic management. We also discovered novel SVs overlapping genes and miRNAs that could impact tumorigenesis. Conclusion: OGM effectively identifies clinically relevant SVs in pediatric brain tumor samples, including SVs that were not discovered by other testing methods. This is a promising technology that allows mapping of the full SV spectrum of pediatric brain tumors. Future clinical studies in larger cohorts will help to determine the effect of SVs on response to treatment/prognosis.

PrgmNr 2775 - The putative tumor suppressor gene*USP13*in familial papillary thyroid carcinoma

View session detail

Author Block: **F. R. Faucz**[1], B. A. Barth[2], A. G. Maria[1], N. Settas[1], R. B. de Alexandre[2], C. A. Stratakis[1], F. Hannah-Shmouni[1]; [1]Section on Endocrinology & Genetics (SEGEN), Eunice Kennedy Shriver Natl. Inst. of Child Hlth.& Human Dev. (NICHD), NIH (NIH), Bethesda, MD, [2]Lab. of Molecular Genetics, Core for Advanced Molecular Investigation (NIMA), Ctr. for Hlth.and Biological Sci. (CCBS), Pontificia Univ.e CatÃ³lica do ParanÃi (PUCPR), Curitiba, Brazil
Disclosure Block: **F.R. Faucz:** None.
Papillary thyroid carcinoma (PTC) is a differentiated type of thyroid malignancy with a high incidence and prevalence. Advances in understanding the genetic basis for this disease revealed the potential involvement of several genes in the formation of thyroid tumors. Recent studies have shown that *USP13*, ubiquitin specific peptidase 13, serves as an important regulator of tumorigenesis and could also work as a tumor suppressor by regulating the PTEN/AKT signaling pathway in some cell types. The aim of this study was to identify genetic variants associated with the development of familial PTC in several members of a non-consagenious Brazilian family, through whole exome sequencing (WES), with *in vitro* characterization. WES analysis was performed in 13 members, including 5 with PTC. Sanger sequencing was used to confirm the main variants identified. A heterozygous missense variant in *USP13* (NM_003940.3) c.1483G>A (p.V495M) fully segregated with the disease. The variant was located in the USP domain of the USP13 protein and highly conserved. Moreover, *USP13* siRNA analysis was performed in MDA-T32 Papillary Thyroid Carcinoma Cells to access the significance of this gene in PTC. MDA-T32 cells treated with *USP13* siRNA showed a decrease of PTEN expression. This analysis suggests that *USP13* may be involved in the pathogenesis of PTC. However, further functional studies are needed to establish the putative role of the *USP13* gene in familial PTC predisposition.

PrgmNr 2776 - Treatment-specific gene regulatory pathways in colorectal cancer patients: CALGB/SWOG 80405 (Alliance)

View session detail

Author Block: **A. Yazdani**[1], H-J. Lenz[2], X. Qu[3], M. M. Bertagnolli[4], A. P. Venook[5], F. Innocenti[1]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]USC Norris Comprehensive Cancer Ctr., Los Angeles, CA, [3]Genentech, San Francisco, CA, [4]Office of the Alliance Group Chair, Brigham and Women's Hosp., Boston, MA, [5]Univ. of California at San Francisco, San Francisco, CA
Disclosure Block: **A. Yazdani:** None.
Cetuximab or bevacizumab combined with chemotherapy are approved regimens for first-line metastatic colorectal cancer. However, hitherto underlying biological pathways can be responsible for the large variation observed in their efficacy and therapeutic responses. To elucidate these mechanisms, we investigated causal gene regulatory pathways related to overall survival (OS) using tumor RNA sequencing and germline SNP data from 1,039 patients with metastatic colorectal cancer treated with cetuximab or bevacizumab through a randomized phase III trial (CALGB/SWOG 80405). In this retrospective study, all analyses have been adjusted for BRAF V600E, all RAS mutations, age, gender, and treatment arm.

We identified 3,161 unique cis-eQTL for tumor gene expression (FDR C7, *CHST14*, *TECPR1*) acting as causal mediators for the impact of germline SNPs on OS. To better understand the regulation of these egenes, we built a transcriptomic-causal network and integrated the results of eQTL and MR analyses into the network. As a result, 7 gene regulatory pathways are identified (FDR Support: U10CA180821, U10CA180882, U24CA196171, https://acknowledgments.alliancefound.org; UG1CA180830, U10CA180888 (SWOG); Lilly, Genentech, and Pfizer; ClinicalTrials.gov Identifier: NCT00265850

PrgmNr 2777 - Trichostatin A sensitivity signature across hematological cell lines

View session detail

Author Block: **B. Przychodzen**; Vanda Pharmaceuticals, DC, DC
Disclosure Block: **B. Przychodzen:** None.
Histone deacetylase inhibitors (HDACi) are small molecules that increase acetylation of lysine residues by blocking histone deactylases. These anticancer agents affect epigenetic and non-epigenetic gene expression resulting in cell cycle arrest of cancer cells. Furthermore HDACi can enhance its anti-tumor effects via the pharmacologic modulation of macrophage (Li et al., 2020). Some HDACiâ□□s such as Trichostatin A (TSA) can also affected the tumor immune microenvironment by suppressing the activity of infiltrating macrophages and inhibiting myeloid-derived suppressor cell recruitment (Li et al.,). We conducted a high throughput screen comparing gene expression profiles in known hematological cell lines to identify transcriptional signatures associated with TSA sensitivity. We selected genes that showed at least 2 fold expression difference and were statistically significant (p

PrgmNr 2778 - A meta-genetic risk score across cardiometabolic risk factors for cardiovascular mortality: a prospective cohort study of UK Biobank

View session detail

Author Block: **S-H. JUNG**[1,2], J-S. Yun[1,3], D. Kim[1,4], H-H. Won[2]; [1]The Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [2]SAIHST, Sungkyunkwan Univ., Samsung Med. Ctr., Seoul, Korea, Republic of, [3]St. Vincentâ□□s Hosp., Coll. of Med., The Catholic Univ. of Korea, Seoul, Korea, Republic of, [4]Inst. for BioMed. Informatics, Univ. of Pennsylvania, Philadelphia, PA
Disclosure Block: **S. Jung:** None.
Background Cardiovascular disease (CVD) is one of the leading causes of mortality worldwide. Thus, early detection of individuals at high risk of CVD is important to lower their risk for cardiovascular (CV) mortality. CV mortality is a complex outcome driven by genetic and environmental factors. Recent studies showed that genetic risk score (GRS) in addition to clinical risk factors has shown the potential for early detect individuals of CVD risk groups. However, the utility of CVD GRS in CV mortality risk group stratification is uncertain. In this study, we generated metaGRS across cardiometabolic (CM) risk factors for CV mortality and evaluated its risk group stratification.
Methods We used data from the UK Biobank with 377,909 European descent participants. We randomly divided the whole dataset into 20% for derivation set and 80% for validation set. The metaGRS generation steps are as follows: i) We obtained GWAS summary statistics for seven well-known CV-related outcomes (stroke, coronary artery disease, chronic kidney disease, heart failure, hypercholesterolemia, hypertension, and type 2 diabetes) and 12 CM-related phenotypes (eGFR, BMI, height, lipid profiles, systolic & diastolic blood pressures, fasting glucose, HbA1c, and smoking status). ii) We generated GRSs for each phenotype and constructed metaGRS using cox proportional hazard model with 10-fold cross validation in the derivation set. The per-GRS hazard ratios (HRs) were applied as the weights in the final model.
Results We categorized participants into four risk subgroups of metaGRS: low (0-19th percentile), intermediate (20-79th percentile), high (80-98th percentile), and very high (99-100th percentile). In the validation set, the best single predictor for CV mortality was CAD GRS which was significantly associated with CV mortality risk (low vs. very high-risk groups; HR=2.96, 95% CI=2.22-3.94). The metaGRS was significantly associated with the highest risk of CV mortality (HR=3.50, 95% CI=2.67-4.60).
Conclusions In summary, metaGRS well stratified risk groups and improved prediction performance of CV mortality risk by combining individual GRSs, suggesting that genetic risk of each CM factor may be partly independent and complementary.

PrgmNr 2779 - Association of a novel locus with rheumatic heart disease in black African individuals: Findings from the RHDGen study

View session detail

Author Block: **T. Machipisa**[1], M. Chong[2], B. Muhamed[3], C. Chishala[3], G. Shaboodien[3], S. Pandie[3], J. de Vries[3], N. Laing[3], A. Joachim[3], R. Daniels[3], M. Ntshekhe[3], C. T. Hugo-Hamman[4], B. Gitura[5], S. Ogendo[5], P. Lwabi[6], E. Okello[7], A. Damasceno[8], C. Novela[8], A. O. Mocumbi[9], G. Madeira[10], J. Musuku[11], A. Mtaja[11], A. ElSayed[12], H. H. Alhassan[12], F. Bode-Thomas[13], B. N. Okeahialam[13], L. J. ZÃ¼hlke[3], N. Mulder[14], R. S. Ramesar[14], M. Lesosky[3], T. Parks[15], H. J. Cordell[16], B. Keavney[17], M. Engel[3], G. Pare[18], RHDGen; [1]Univ. of Cape Town (Home institution)/McMaster Univ., Population Hlth.Res. Inst., Genetic Molecular Epidemiology Lab (visiting PhD Fellow), Cape Town, South Africa, [2]Population Hlth.Res. Inst. (Genetic Molecular Epidemiology Lab), Hamilton, ON, Canada, [3]Univ. of Cape Town, Cape Town, South Africa, [4]Rheumatic Heart Disease Clinic, Windhoek Central Hosp., Ministry of Hlth.and Social Services, Windhoek, Namibia, [5]Cardiology Dept. of Med., Kenyatta Natl. Hosp., Univ. of Nairobi, Nairobi, Kenya, [6]Uganda Heart Inst., Dept.s of Adult and Pediatric Cardiology, Kampala, Uganda, [7]Uganda Heart Inst., Dept.s of Adult and Pediatric Cardiology, Uganda, Uganda, [8]Faculty of Med., Eduardo Mondlane Univ. / Nucleo de InvestigaÃ§ao, Departamento de Med., Hosp. Central de Maputo, Maputo, Mozambique, [9]Inst. Natl. de SaÃºde MinistÃ©rio da SaÃºde â MoÃ§ambique, Maputo, Mozambique, [10]WHO Mozambique, CP 377 Rua Pereira Marinho 280, Maputo, Mozambique, [11]Univ. Teaching Hosp. â Childrenâs Hosp., Univ. of Zambia, Lusaka, Zambia, [12]Dept. of Cardiothoracic Surgery, Alshaab Teaching Hosp., Alazhari Hlth.Res. Ctr., Alzaiem Alazhari Univ., Khartoum, Sudan, [13]Dept.s of Paediatrics and Med., Jos Univ. Teaching Hosp. and Univ. of Jos, Jos, Nigeria, [14]Univ Cape Town, Cape Town, South Africa, [15]Wellcome Ctr. for Human Genetics, Univ. of Oxford, Oxford, OX3 7BN, Oxford, United Kingdom, [16]Newcastle Univ., Newcastle upon Tyne, United Kingdom, [17]The Univ. of Manchester & the Manchester Univ. NHS Fndn. Trust, Manchester, United Kingdom, [18]Hamilton, ON, Canada

Disclosure Block: **T. Machipisa:** None.

IMPORTANCE: Rheumatic heart disease (RHD), a sequela of rheumatic fever characterized by permanent heart valve damage, is the leading cause of cardiac surgery in Africa. However, its pathophysiologic characteristics and genetics are poorly understood. Understanding genetic susceptibility may aid in prevention, control, and interventions to eliminate RHD. OBJECTIVE: To identify common genetic loci associated with RHD susceptibility in Black African individuals. DESIGN, SETTING, AND PARTICIPANTS: This multicenter case-control genome-wide association study (GWAS), the Genetics of Rheumatic Heart Disease, examined more than 7 million genotyped and imputed single-nucleotide variations. The 4809 GWAS participants and 116 independent trio families were enrolled from 8 African countries between December 31, 2012 and March 31, 2018. All GWAS participants and trio probands were screened by use of echocardiography. Data analyses took place from May 15, 2017, until March 14, 2021. MAIN OUTCOMES AND MEASURES: Genetic associations with RHD. RESULTS: This study included 4809 African participants (2548 RHD cases and 2261 controls; 3301 women [69%]; mean [SD] age, 36.5 [16.3] years). The GWAS identified a single RHD risk locus, 11q24.1 (rs1219406 [odds ratio, 1.65; 95%CI, 1.48-1.82; P = 4.36 Ã~ 10â8]), which reached genome-wide significance in Black African individuals. Our meta-analysis of Black (n = 3179) and admixed (n = 1055) African individuals revealed several suggestive loci. The study also replicated a previously reported association in Pacific Islander individuals (rs11846409) at the immunoglobulin heavy chain locus, in the meta-analysis of Black and admixed African individuals (odds ratio, 1.16; 95%CI, 1.06-1.27; P = 1.19 Ã~ 10â3). The HLA (rs9272622) associations reported in Aboriginal Australian individuals could not be replicated. In support of the known polygenic architecture for RHD, over-transmission of a polygenic risk score from unaffected parents to affected probands was observed (polygenic transmission disequilibrium testing mean [SE], 0.27 [0.16] SDs; P = .04996), and

the chip-based heritability was estimated to be high at 0.49 (SE = 0.12; P = 3.28 Ã~ 10â□□5) in Black African individuals. CONCLUSIONS AND RELEVANCE: This study revealed a novel candidate susceptibility locus exclusive to Black African individuals and an important heritable component to RHD susceptibility in African individuals.

PrgmNr 2780 - Cardiac enhancer variants at the *MTSS1* locus are associated with smaller, more contractile hearts

View session detail

Author Block: **M. BURKE**[1], M. Morley[1], Y. Yang[1], T. Drivas[2], D. Choe[1], S. Day[1], K. Musunuru[1], M. Li[3], M. D. Ritchie[4], T. P. Cappola[1]; [1]Univ of PA, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Univ Pennsylvania, Philadelphia, PA, [4]Univ. of Pennsylvania, Philadelphia, PA

Disclosure Block: **M. Burke:** None.

Introduction: Common variants at the *MTSS1* locus influence cardiac structure, function, and failure in multiple genome wide association studies (GWAS), yet the genetic mechanisms responsible remain unclear. We previously identified common variants residing in a cardiac specific enhancer in *cis* to *MTSS1* that influence MTSS1 transcript abundance in the human left ventricle (LV). Here we used bioinformatic approaches to determine if these enhancer variants explain the association between common *MTSS1* variation and multiple cardiac traits. We hypothesized that enhancer alleles that reduce *MTSS1* expression would be associated with smaller, more contractile hearts.

Methods: To confirm a shared causal variant between MTSS1 expression in the LV and cardiac traits, we combined human expression quantitative trait locus data from GTEx with GWAS data from the UK Biobank (UKBB) using co-localization. GTEx consists of 386 samples with RNAseq expression data from the LV. The UKBB is a prospective cohort of over 500,000 patients ages 40-69 recruited between 2006 and 2010 across the United Kingdom with extensive available phenotype data. In addition to baseline questionnaire and health assessment, electronic health record, and biomarker data, genotyping has been undertaken for the entire cohort. We utilized quantified cardiac MRI data from returned dataset 2383, which provides 82 cardiac phenotypes in 26,892 individuals. We obtained the relevant UKBB data as part of approved protocol 65965. Genome association tests were performed at the *MTSS1* locus (chr8:124550790 - 124728429) using Plink v.2.0; co-localization analysis was performed using Coloc.

Results: Co-localization analysis of 26,892 patients enrolled in the UKBB who underwent cardiac MRI imaging revealed three overlapping variants rs12541595, rs35006907, and rs34866937 in the *MTSS1* enhancer region significantly associated with LV expression of *MTSS1* as well as with multiple cardiac MRI traits. Notably, these traits included LV ejection fraction (LVEF) ($p = 1.3 \times 10^{-10}$), LV end-systolic volume ($p = 1.2 \times 10^{-8}$), and global circumferential strain ($p = 1.1 \times 10^{-7}$). Variants in this enhancer region that decrease *MTSS1* expression are associated with decreased volumes, increased LVEF, and improved global circumferential strain, all of which are consistent with a cardioprotective phenotypic effect.

Conclusions: In this study, we utilize a bioinformatics approach to clarify the genetic mechanisms by which a previously identified locus modulates cardiac traits. These results clearly support a cardioprotective phenotype associated with reduction in *MTSS1* gene expression in the LV.

PrgmNr 2781 - Cell-type-specific gene expression and transcriptional networks reveal *Adamts2* as a robust regulator of cardiac homeostasis during heart failure

View session detail

Author Block: **C. Rau**[1], C. Lahue[1], S. Lee[1], D. J. Chapski[2], M. Rosa-Garrido[2], S. Ren[2], T. M. Vondriska[3], Y. Wang[3]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Univ. of California Los Angeles, Los Angeles, CA, [3]Univ. of California Los Angeles, Los Angeles, NC
Disclosure Block: **C. Rau:** None.
Heart failure (HF) is a highly heterogeneous disorder characterized by the interactions of multiple genetic and environmental factors as well as the interaction of different cell types in the heart. Although reductionistic approaches have successfully identified many genes involved in HF, heritability studies show many more candidate genes than can be studied individually. By utilizing cell-type-specific gene expression paired with additional transcriptomic data from a large cohort of mice, we sought to identify important drivers of HF using a systems genetics approach. Mice from 93 unique inbred lines of the Hybrid Mouse Diversity Panel were given 30 ug/g/day of isoproterenol for three weeks via osmotic minipump to induce heart failure. Transcriptomes were generated from these mice and the weighted Maximal Information Component Analysis (wMICA) algorithm was applied to generate transcriptomic gene networks. Cardiomyocytes and Fibroblasts were isolated from both control and isoproterenol-treated adult C57BL/6J hearts using a Langendorff apparatus (n=3 per sex/treatment) and transcriptomes were generated. Significantly differentially expressed genes were identified using DESEQ2 and used to query the wMICA-derived network, identifying the gene *Adamts2* as a potential regulator of cardiac hypertrophy. Follow-up *in vitro* and *in vivo* work has demonstrated that *Adamts2* knockdown significantly blunts the hypertrophic effect of isoproterenol on cardiomyocytes while simultaneously reducing fibroblast proliferation and increasing apoptosis as measured by TUNEL staining. Careful examination of the gene network reveals evidence of paracrine signaling between cardiomyocytes and fibroblasts and suggests a key trans-cell-type role of *Adamts2* in the regulation of HF after catecholamine stimulation. In conclusion, Co-expression network algorithms combined with cell-type-specific transcriptomics identified *Adamts2* as a driver of HF. *Adamts2* plays an important role via paracrine signaling in the proliferative response of fibroblasts and hypertrophic response of cardiomyocytes to catecholamines. Further mechanistic analysis of *Adamts2* will further reveal its role in the progression of heart failure.

PrgmNr 2782 - Dissecting 90 percent of lipoprotein(a) heritability in 487,571 UK Biobank participants

View session detail

Author Block: **R. E. Mukamel**[1], R. E. Handsaker[2], M. A. Sherman[3], A. Barton[4], Y. Zheng[1], S. A. McCarroll[4], P-R. Loh[5]; [1]Brigham and Women's Hosp./Harvard Med. Sch., Boston, MA, [2]Charlemont, MA, [3]Massachusetts Inst. of Technology, Marblehead, MA, [4]Harvard Med. Sch., Boston, MA, [5]Brigham and Women's Hosp. / Harvard Med. Sch., Boston, MA
Disclosure Block: **R.E. Mukamel:** None.
Lipoprotein(a) concentration [Lp(a)] is a highly heritable (h^2>90%), monogenic cardiovascular risk factor whose primary genetic determinant is the size of the *LPA* kringle IV-2 (KIV-2) repeat, which explains ~60% of Lp(a) heritability. However, the full spectrum of other functional sequence variants in *LPA* and the way in which they shape Lp(a) variation has been unclear due to the difficulty of genotyping KIV-2 variation from high-throughput genomic assays. We developed methods to accurately estimate haplotype-resolved KIV-2 allele lengths in N=49,959 exome-sequenced UK Biobank (UKB) participants (RMSE ~1 repeat unit), and to impute KIV-2 lengths into SNP haplotypes for the remaining N=437,612 UKB participants (imputation R^2=0.9).
To systematically identify and measure the effects of additional Lp(a)-altering variants, we isolated the contributions of individual *LPA* haplotypes by analyzing heterozygous carriers of two coding variants known to create nonfunctional alleles (rs41272114 and rs41259144, combined MAF=0.05). This approach created an effective haploid model for Lp(a). Stepwise conditional analyses controlling for KIV-2 length pointed to a series of 20 additional protein-altering and 5â UTR variants that appeared to further shape Lp(a). Together with KIV-2 length, these variants explained 90% of heritable Lp(a) variance in European-ancestry participants in a model that accounted for nonlinear and cis-epistatic effects. These variants also largely explained the ~4-fold variation in median Lp(a) across populations: the higher Lp(a) commonly observed in African populations appeared to result from a relative paucity of alleles carrying a large-effect Lp(a)-reducing coding or splice-altering SNP in LPA (affecting 13% vs. 43% for African vs. European alleles) and the high frequency of the Lp(a)-increasing 5â UTR variant rs1800769 (46% vs. 17% for African vs. European alleles).
The accuracy of genetically predicted Lp(a) also enabled insights into epidemiological associations with Lp(a). The myocardial infarction risk-increasing effect of higher Lp(a) extended to extreme Lp(a) levels in UKB (OR=3.1, 95% CI=1.9-5.2 for individuals with predicted Lp(a)>400 nmol/L). In contrast, lower genetically predicted Lp(a) did not associate with increased type-2 diabetes (T2D) risk, suggesting that the 17% (s.e. 1%) lower levels of measured Lp(a) observed in T2D patients represents reverse causation resulting from T2D itself, T2D-related liver comorbidities, or T2D medication.
These results demonstrate that complex genetics can underlie even monogenic traits and inform genetic approaches to cardiovascular risk stratification.

PrgmNr 2783 - Genetic association analysis of dilated cardiomyopathy in the VA Million Veteran Program

View session detail

Author Block: **J. E. Huffman**[1,2], B. Charest[1], J. Gaziano[1,2,3], P. W. F. Wilson[4,5], L. Phillips[4,5], J. Joseph[1,2,3], Y. Sun[5,4], K. Aragam[6,1,7], VA Million Veteran Program; [1]VA Boston Hlth.care System, Jamaica Plain, MA, [2]Harvard Med. Sch., Boston, MA, [3]Brigham and Women's Hosp., Boston, MA, [4]Atlanta VA Med. Ctr., Decatur, GA, [5]Emory Univ., Atlanta, GA, [6]Massachusetts Gen. Hosp., Boston, MA, [7]Broad Inst., Cambridge, MA

Disclosure Block: **J.E. Huffman:** None.

Background/Aims: Dilated cardiomyopathy (DCM) is a major cause of heart failure and the leading global cause of heart transplantation. Rare, Mendelian drivers of DCM are well-established, but the potential contribution of common genetic variation to DCM risk has only recently begun to emerge. To date, common variant association studies (CVAS) of DCM have relied predominantly on prospectively recruited cases of DCM. We sought to conduct a CVAS of DCM in a large biobank. **Methods:** We established an electronic health record-based phenotypic definition of DCM in the Million Veteran Program (MVP), using an ICD-10 code of I42.0 to define cases, and excluding heart failure cases and other cardiomyopathies from the controls. We performed ancestry-specific CVAS then trans-ethnic meta-analysis in MVP participants of European and African genetic ancestry. We pursued replication of genomic findings in the UK Biobank. We also assessed the association of sentinel DCM SNPs with hypertrophic cardiomyopathy (HCM) in MVP, and subclinical cardiac magnetic resonance imaging (MRI) measures of left ventricular (LV) structure and function in UK Biobank. **Results:** Among 459,937 MVP study participants, mean age was 61.5, 43,971 (9.6%) were female, and 5,597 met phenotypic criteria for DCM (N=3,786 European ancestry, N=1,811 African ancestry). In the trans-ethnic meta-analysis, we identified eight genomic regions associated with DCM at genome-wide significance (P -8). These included three of four known DCM associations near *HSPB7*, *LSM3*, and *BAG3*; and five novel genomic regions near *SLC39A8*, *CDKN1A*, *CD36*, *NEDD4L*, and *MAP3K7CL*. Of the five novel genomic regions, three (*CDKN1A*, *NEDD4L*, and *MAP3K7CL*) were replicated in the UK Biobank (all p-4; N=899 DCM cases, 449,574 controls). While only four of the significant DCM loci achieved nominal significance with HCM (N=1,533 HCM cases, 465,877 controls), we observed a previously-reported trend that the direction of effect for sentinel SNPs was opposite between these two cardiomyopathies. SNPs associated with increased DCM risk also associated with greater LV size and reduced LV function - the defining traits of DCM - even in UK Biobank participants without clinical heart failure or cardiomyopathy (N > 31,614). **Conclusions:** We identified 5 novel, common genetic loci associated with DCM and replicated three of these in an external DCM cohort, and provided plausible supporting evidence through association with DCM-relevant subclinical MRI traits. Our findings highlight the potential to leverage large, multi-ethnic biobank populations to enhance common variant discovery for DCM.

PrgmNr 2784 - Genetic variation in/near *LIPC*, *MGAT1*, and *APOA1* is associated with longitudinal change in serum lipid levels in Samoan adults

View session detail

Author Block: **S. Liu**[1], H. Fu[2], N. L. Hawley[3], J. C. Carlson[2,1], A. Pomer[4], E. M. Russell[1], H. Cheng[5], T. Naseri[6,7], M. S. Reupena[8], R. Deka[5], C. C. Choy[7], E. E. Kershaw[9], S. T. McGarvey[7,10], R. L. Minster[1], D. E. Weeks[1,2]; [1]Dept. of Human Genetics, Graduate Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [2]Dept. of Biostatistics, Graduate Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [3]Dept. of Chronic Disease Epidemiology, Sch. of Publ. Hlth., Yale Univ., New Haven, CT, [4]Dept. of Anthropology, Yale Univ., New Haven, CT, [5]Dept. of Environmental and Publ. Hlth.Sci., Coll. of Med., Univ. of Cincinnati, Cincinnati, OH, [6]Ministry of Hlth., Government of Samoa, Apia, Samoa, [7]Intl. Hlth.Inst., Dept. of Epidemiology, Sch. of Publ. Hlth., Brown Univ., Providence, RI, [8]Lutia i Puava ae Mapu i Fagalele, Apia, Samoa, [9]Div. of Endocrinology and Metabolism, Dept. of Med., Univ. of Pittsburgh, Pittsburgh, PA, [10]Dept. of Anthropology, Brown Univ., Providence, RI

Disclosure Block: **S. Liu:** None.

Samoan adults have a high incidence of cardiovascular disease which is significantly associated with abnormal lipid levels. Here we studied the relationship between longitudinal changes in fasting serum lipid levels (high-density lipoprotein cholesterol (HDL), low density lipoprotein cholesterol (LDL), triglycerides (TG) and total cholesterol (TC)) and eight SNPs that were associated with cross-sectional lipid levels in previous genome-wide association scans in Samoan adults. We focused on 519 adult participants who completed assessments in both 2010 and 2017-2019. We constructed rate-of-change models for 32 trait-marker pairs to evaluate the relationship between longitudinal changes in lipid levels and each SNP using linear regression adjusting for age, ageÂ², sex, and other baseline covariates selected by LASSO which was performed in 2,304 Samoans from the 2010 study that do not overlap with the 519 participants studied here. Selected covariates varied by trait, but often included census region, smoking status, diabetes status, waist-hip ratio, and arm circumference. We found that the A allele of rs10438284 near *LIPC* is associated with lower rate-of-change in HDL (Î² = â□□0.179 mg/dL per year, p = 0.010). The T allele of rs1038143 near *MGAT1* was associated with higher rate-of-change in LDL (Î² = 0.470 mg/dL per year, p = 0.048). Residential census region was also associated with rate-of-change in LDL with more rural areas having higher rates-of-change compared to the Apia Urban Area (Î² = 1.079 mg/dL per year for peri-urban Northwest Upolu, p = 0.008; Î² = 1.103 mg/dL per year for rural Rest of Upolu, p = 0.009). We also observed a positive association between the C allele of rs964184 near *APOA1* and rate-of-change in triglycerides (Î² = 1.17 mg/dL per year, p = 0.026). Our study showed that genetic variation in/near *LIPC*, *MGAT1,* and *APOA1* as well as demographic factors have an impact on longitudinal changes in serum lipid levels. A better understanding of these factors might inform who is at greatest risk of achieving unhealthy serum lipid levels, lead to future interventions, and help reduce the burden of cardiovascular disease in Samoans.

PrgmNr 2785 - Genome-wide association study of coronary heart disease in chronic kidney disease: the Chronic Renal Insufficiency Cohort (CRIC) study

View session detail

Author Block: **J. Wen**[1], B. Lin[2], G. Li[3], Q. Sun[1], Y. Li[4], N. Franceschini[5]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Univ. of North Carolina, Chapel Hill, NC, [3]Univ. of North Carolina at Chapel Hill, BELLEVUE, WA, [4]Univ North Carolina, Chapel Hill, NC, [5]Univ North Carolina at Chapel Hill, Chapel Hill, NC

Disclosure Block: **J. Wen:** None.

Chronic kidney disease (CKD) is associated with premature and widespread atherosclerosis, and atherosclerotic cardiovascular disease is a main cause of death in CKD patients. The underlying genetic factors for increased cardiovascular disease susceptibility in CKD are unknown. To identify loci associated with coronary heart disease in CKD, we performed a genome-wide association study (GWAS) of incident myocardium infarction in the CRIC study, a multi-center cohort of CKD participants recruited from seven U.S. clinical centers. Genotypes were imputed to TOPMed whole genome sequencing multi-ethnic reference panel (freeze 5b). We performed time to event analysis using additive models adjusted for age, sex, site of recruitment, baseline estimated glomerular filtration rate (eGFR) and genotype principal components. Among 2,793 participants without coronary heart disease at baseline (mean age 56.4 years, 55% men, 42% African Americans, 58% whites, mean eGFR 45.3 ml/min/1.73 m^2), 214 developed a clinical event at an average of 9.1 years of follow-up. GWAS (with genomic control Î»=1.003) identified two loci at the genome-wide significance threshold (PMYO6) and an intronic variant on chromosome 9 (*TRPM6*). Both of these are novel loci comparing to the GWAS catalog loci for coronary heart disease in non-CKD general population. Further checking gene expression in different tissues from GTEx v8, we find *MYO6* is highly expressed in kidney and brain; *TRPM6* expressed in brain. We also replicated, accounting from multiple testing, 36 known coronary heart disease loci including the *APOB*, *LPL*, *FGF5*, *EDNRA* and *CDKN2B* loci. These findings suggest that both shared and CKD-specific genetic factors contribute to clinical coronary heart disease in individuals with CKD.

PrgmNr 2786 - Leveraging a founder population to identify novel rare-population genetic determinants of lipidome

View session detail

Author Block: **M. E. Montasser**[1], S. Aslibekyan[2], V. Srinivasasainagendra[2], H. K. Tiwari[2], A. Patki[2], M. Bagheri[3], T. Kind[4], D. Barupal[4], S. Fan[4], J. A. Perry[1], K. Ryan[5], D. K. Arnett[6], A. Beitelshees[1], M. Irvin[2], J. O'Connell[1]; [1]Univ. of Maryland, Baltimore, MD, [2]Univ. of Alabama, Birmingham, AL, [3]Vanderbilt Univ., Nashville, TN, [4]West Coast Metabolomics Ctr., Davis, CA, [5]Univ. of Maryland, Baltimore, Baltimore, MD, [6]Univ. of Kentucky, Lexington, KY

Disclosure Block: **M.E. Montasser:** Grant/Contracted Research Support (External); Regeneron Pharmaceuticals Inc..

Identifying the genetic determinants of variation in lipid species (lipidome) may provide deeper mechanistic insight in CVD risk beyond traditional lipids. Previous studies have been largely population based and thus only powered to discover common genetic variants. Founder populations represent a powerful resource to accelerate discovery of novel biology associated with rare population alleles that have risen to higher frequency by genetic drift. We performed a GWAS of 355 lipid species in 650 individuals from the Amish founder population. The population-based Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study was used for replication and fine mapping along with publicly available functional annotations. We report for the first time the lipid species associated with two rare-population but Amish-enriched lipid variants: the cardioprotective *APOC3*_rs76353203 (MAF=0.02 vs 0.0008 in the general European population) was associated with lower levels of 3 phosphatidylethanolamines (PE(36:2), (38:6), (34:2)) (p =6.3E-13 to 1.3E-10), and the familial hypercholesterolemia variant *APOB*_rs5742904 (MAF=0.06 vs 0.0004) was associated with increased levels of several cholesterol esters, sphingolipids and phospholipids (p =7.9E-25 to 3.1E-10). We also identified novel associations for 3 rare-population Amish-enriched loci: The first is a missense variant (rs536055318, A263T) (MAF=0.07 vs 0.001) in an active transcription start site in the promoter of the *GLTPD2* gene associated with lower sphingomyelins (SM(d40:0)) (p =1.1E-12). *GLTPD2* is mainly expressed in liver and kidney and plays a role in the intermembrane transfer of glycolipids. The second is a potentially disease-causing splice donor missense variant (rs771033566, Val344Leu) (MAF=0.04 vs 0.01) in the *CERS5* gene, associated with lower SM(32:2) (p =2.2E-14). *CERS5* is one of the six members of the ceramide synthase gene family which plays a major role in the sphingolipid metabolic salvage pathway. The third is an intronic variant (rs531892793) (MAF=0.04 vs 0.0001) in the *AKNA* gene associated with lower levels of 5 glucosylceramide species (GlcCer(d38:1), (d40:1), (d41:1), (d42:1), (d42:2)) (p =4.0E-17 to 1.9E-12). rs531892793 is located in the promoter flanking region and has the top ENCODE DNase score of 1000 indicating very strong evidence of being a DNase hypersensitivity site, an eigenPC score of 3.5 indicating a strong functional prediction based on conservation and allele frequency, and predicted to affect transcriptional factor. Our results show the power of founder populations to discover novel biology associated with rare population variants through allele enrichment.

PrgmNr 2787 - Multi-trait GWAS of atherosclerosis detects novel loci and potential therapeutic targets

View session detail

Author Block: **W. P. Bone**[1], T. R. Bellomo[2], B. Y. Chen[1], K. Gawronski[1], D. Zhang[1], J. Park[3], M. Levin[4], N. Tsao[4], D. Klarin[5], J. Lynch[6], T. L. Assimes[7], M. Gaziano[8], P. W. Wilson[9], K. Cho[10], M. Vujkovic[1], the VA Million Veteran Program, C. J. O'Donnell[10], K-M. Chang[4], P. TSAO[11], D. J. Rader[12], M. D. Ritchie[1], S. M. Damrauer[13], B. F. Voight[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [3]Perelman Sch. of Med. at the Univ. of Pennsylvania, Philadelphia, PA, [4]Corporal Michael J. Crescenz VA Med. Ctr., Philadelphia, PA, [5]Massachusetts Gen. Hosp., Boston, MA, [6]Salt Lake City VA Med. Ctr., Bedford, MA, [7]Stanford Univ Sch Med., Stanford, CA, [8]VA Boston Hlth.care System, Jamaica Plain, MA, [9]Atlanta VA Med. Ctr., Decatur, GA, [10]VA Boston Hlth.care System, Boston, MA, [11]VAPAHCS, Palo Alto, CA, [12]Univ of Pennsylvania, Philadelphia, PA, [13]Hosp. of the Univ. of Pennsylvania, Philadelphia, PA

Disclosure Block: **W.P. Bone:** None.

Atherosclerosis, which is the narrowing of the arterial walls via accumulation of cholesterol-rich arterial plaques, is the leading cause of vascular disease worldwide, including myocardial infarction and ischemic stroke. Although atherosclerosis affects arteries throughout the body, previous genome-wide association studies (GWAS) have been performed on specific atherosclerotic phenotypes such as coronary artery disease (CAD) and peripheral artery disease (PAD). There is substantial evidence to suggest that these more specific atherosclerosis phenotypes share a common genetic etiology. We performed a series of multi-trait GWAS using combinations of two atherosclerosis traits and seven atherosclerosis risk factor traits and detected 31 novel pleiotropic loci. We performed these multi-trait GWAS using the N-GWAMA multi-trait GWAS method and summary statistics for CAD (van der Harst et al. 2018), PAD (Klarin et al. 2019), body mass index (Pulit et al. 2019), type II diabetes (Vujkovic et al. 2020), smoking initiation (Wootton et al. 2020), and lipid traits (Klarin et al. 2019). We identified candidate causal genes for 14 of these loci through colocalization analysis with GTEx expression quantitative trait locus (eQTL) data. *VDAC2* and *PCSK6* are two candidate causal genes that our results and previous literature suggest are potential therapeutic targets. *VDAC2* eQTLs in aorta and tibial artery colocalized with a multi-trait GWAS signal detected in the CAD PAD multi-trait GWAS. Previous work has shown that *VDAC2* regulates apoptosis, and our results suggest increased *VDAC2* expression in smooth muscle cells could increase smooth muscle cell accumulation in atherosclerotic plaques. A sQTL (splicing QTL) for *PCSK6* in liver colocalized with a multi-trait GWAS signal between PAD and LDL. Further analysis of the sQTL signal suggested that the effect allele correlates with a more active isoform of *PCSK6*, which could increase lipid fractions and risk of atherosclerosis. These results show that joint analysis of atherosclerotic disease traits and their risk factors allows for identification of unified biology that may offer the opportunity for therapeutic manipulation.

PrgmNr 2788 - Prioritizing genes from genome-wide meta-analyses of blood lipid levels

View session detail

Author Block: **Y. Wang**[1], S. E. Graham[2], S. Kanoni[3,4], S. Ramdas[5], K. F. Bhatti[3], I. Surakka[2], P. Natarajan[6,7,8,9], Y. V. Sun[10,11,12], T. L. Assimes[13,14], P. Deloukas[3,4], C. D. Brown[5], C. J. Willer[2,15,16], G. M. Peloso[1], Million Veteran Program, Global Lipids Genetics Consortium; [1]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [2]Dept. of Internal Med., Div. of Cardiovascular Med., Univ. of Michigan, Ann Arbor, MI, [3]Clinical Pharmacology, William Harvey Res. Inst., Barts and the London Med. Sch., Queen Mary Univ. of London, London, United Kingdom, [4]Ctr. for Genomic Hlth., Queen Mary Univ. of London, London, United Kingdom, [5]Dept. of Genetics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [6]Cardiology Div., Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA, [7]Dept. of Med., Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA, [8]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, [9]Cardiovascular Res. Ctr. and Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, Boston, MA, [10]VA Atlanta Hlth.care System, Decatur, GA, [11]Dept. of Epidemiology, Emory Rollins Sch. of Publ. Hlth., Atlanta, GA, [12]Dept. of BioMed. Informatics, Emory Sch. of Med., Atlanta, GA, [13]Dept. of Med., Div. of Cardiovascular Med., Stanford Univ., Stanford, CA, [14]VA Palo Alto Hlth.Care System, Palo Alto, CA, [15]Dept. of Computational Med. and Bioinformatics, Univ. of Michigan, Ann Arbor, MI, [16]Dept. of Human Genetics, Univ. of Michigan, Ann Arbor, MI

Disclosure Block: **Y. Wang:** None.

The Global Lipids Genetics Consortium trans-ancestry genome-wide association study (GWAS) of 1.65 million individuals has identified 2,399 index variants associated with any of 5 lipid phenotypes (LDL-C, HDL-C, total-cholesterol, triglyceride, and nonHDL-C). A limitation of GWAS is that the causal gene for associated variants is not readily discernible. We aimed to identify the likely functional genes for our lipid associated index variants by employing four locus-dependent approaches: 1) the closest gene to the index variant, 2) genes with associated protein coding variants, 3) expression QTLs (eQTLs) colocalized genes, and 4) genes prioritized by transcript-wide association studies (TWAS), and two genome-wide dependent approaches: 1) gene-level Polygenic Priority Score (PoPS), and 2) co-regulation of gene expression and reconstituted gene sets (DEPICT). We further quantified how often the same gene was prioritized by multiple methods for each index variant and determined scores that ranged from 1-6, based on the number of methods that prioritized the gene. To evaluate the performance of each prioritization method and the credibility of the scores, we used 23 genes known to cause Mendelian dyslipidemias (gold standard set) and 747 mouse knockout genes causing lipid phenotypes (silver standard set). As validation, we estimated the proportion of genes in the gold and silver standard sets that were correctly prioritized based on our selection of approaches. Within the gold standard set, the proportion of genes correctly prioritized were: 79.6% by TWAS, 78.6% by PoPS, 60.2% by DEPICT, 57.3% by protein coding variants, and 42.7% by the closest gene. A similar rank order for the silver standard genes was observed. However, since TWAS had prioritized 3,511 genes which was greater than other methods (~1000), it had a much smaller proportion of correctly prioritized genes out of the total number of genes prioritized by those methods. Based on the above, we assigned candidate causal genes for each index variant with an *ad hoc* confidence level: high, medium high, medium low and low. In addition to the 23 gold-standard genes, we prioritized a further 94 genes with high confidence. Text mining of PubMed titles and abstracts suggests that 43.6% of the 94 high confidence genes have minimal (

PrgmNr 2789 - Proteomic analysis of 92 circulating proteins and their effects in cardio-metabolic diseases

View session detail

Author Block: **C. Carland**[1], D. Zanetti[2], G. Png[3], A. Gilly[3], E. Zeggini[3], T. L. Assimes[4], SCALLOP Consortium; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]CVI Stanford Cardiovascular Inst., Palo Alto, CA, [3]Helmholtz Zentrum MÃ¼nchen, Neuherberg, Germany, [4]Stanford Univ Sch Med., Stanford, CA
Disclosure Block: **C. Carland:** None.
Introduction Human plasma contains many circulating proteins derived from multiple organs and that participate in a wide range of biological processes. Circulating proteins play an important clinical role as biomarkers (eg, NT-proBNP) and also as drug targets (eg, PCSK9). Recent advances in high throughput quantification have allowed for the scale up of investigating association of genetic variants with circulating protein levels.
Methods We conducted genome-wide association studies of autosomal chromosomes in 12 cohorts involving up to 22,997 European ancestry individuals to identify protein quantitative trait loci (pQTLs) for 92 plasma proteins on the â□□metabolismâ□□ panel using OLINKâ□□s proximity extension assay. Meta-analysis was conducted with a random effects model using GWAMA and independent associations identified with PLINK and GCTA-COJO. Heritability and pairwise genetic correlations were calculated using LD score regression. Additionally, we performed a sex-stratified meta-analysis on up to 10,870 men and 12,127 women. Two sample Mendelian randomization (MR) was conducted using TwoSampleMR package on R.
Results We identified 503 significant conditionally independent pQTLs (337 cis and 166 trans) in 77 proteins at P 0.25, >0.5, and >0.75. A sex stratified analysis revealed a total of 552 pQTLs in women and 258 pQTLs in men (p **Conclusion** Large-scale genome wide meta analyses across circulating plasma proteins provides insight into causal mechanisms of disease and potential drug therapeutic targets. Sex stratified analysis suggests similar genetic underpinnings of circulating plasma proteins, however a few sex-specific pQTLs may point to some differences.

PrgmNr 2790 - Rare variant carrier status on risk of coronary artery disease among individuals with high polygenic risk

View session detail

Author Block: **M. Shivakumar**[1], S-H. JUNG[2], J-S. Yun[3], M. D. Ritchie[2,4], D. Kim[2]; [1]Dept. of Biostatistics, Epidemiology and Informatics, Univ. of Pennsylvania, Philadelphia, PA, USA, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]The Catholic Univ. of Korea, Bryn Mawr, PA, [4]Inst. for BioMed. Informatics, Univ. of Pennsylvania, Philadelphia, PA
Disclosure Block: **M. Shivakumar:** None.

Polygenic risk scores (PRS) derived from large GWAS studies have been able to capture a large portion of genetic risk for coronary artery disease (CAD) risk. Previous studies have shown that the PRS for CAD (PRS_{CAD}) can be used to identify the subgroup from the population that is at greater than three-fold increased risk for CAD. However, in general, rare variants are known to have larger impacts than common variants, and complex diseases like CAD are influenced by both common and rare genetic variants. In this study, we examined the changes in the risk conferred by carriers of rare variants in known genes associated with CAD.

Among 154,799 unrelated individuals of European ancestry, we identified 2,553 CAD cases with both genotype and exome sequencing data available in UK Biobank (UKBB). PRS_{CAD} for all samples were calculated using genotype data using LDpred and summary statistics from the CARDIoGRAMplusC4D GWAS study. The rare variants were obtained from the exome sequencing data and were filtered to keep only predicted loss of function (LOF) variants, variants predicated with high impact by Variant Effect Predictor, and missense variants with REVEL score > 0.5. We split samples into 10 groups based on 5 PRS_{CAD} quantiles and rare variant carrier status. We calculated the risk of CAD between the first and last quantile combined with rare variant carrier status.

Of the 47 known replicated genes for CAD, we found 22 genes where the carriers of rare variants showed higher odds of CAD risk than non-carriers, and in the other 25 genes, the carriers of rare variants showed lower odds of CAD risk than the non-carriers in the high risk PRS group. The samples with LOF rare variants in genes such as *MIA3* (Lowest vs. highest PRS with non-carrier OR: 2.00, carrier OR: 4.13) and *TCF21* (non-carrier OR: 2.00, carrier OR: 3.09) showed higher OR. The high expression of MIA3 is expected to promote atheroprotective vascular smooth muscle cells and higher expression *TCF21* is known to inhibit risk for coronary artery disease. On the contrary, genes like *SMAD3* (non-carrier OR: 2.02, carrier OR: 1.25E-05) and HHIPL1 (non-carrier OR: 2.01, carrier OR: 1.29E-05) showed lower odds, and their expression is known to increase CAD progression.

We showed that the LOF rare variant carriers in the high risk PRS groups may increase or decrease the odds of CAD risk. This study also shows the need to take rare variants into consideration for risk stratification. Robust methods for integration of rare variants with PRS are greatly needed to further improve risk prediction for CAD and allow for the development of improved preventive measures and also novel therapeutic opportunities for CAD.

PrgmNr 2791 - *OCRL* tissue-specific alternative splicing in humans results in divergent phenotypes

View session detail

Author Block: **G. IANNELLO**[1], C. Sena[1], L. Pais[2], R. Sathanayagam[3], C. Genetti[4], A. Noel[5], C. Phornphutkul[6], L. Benson[4], K. Dannheim[7], P. B. Agrawal[8,9], G. Stratigopoulos[1], V. Thaker[1]; [1]Div. of Molecular Genetics, Dept. of Pediatrics, Columbia Univ. Irving Med. Ctr., New York, NY, [2]Broad Inst. Ctr. for Mendelian Genomics and Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [3]SUNY Downstate Hlth.Sci. Univ., New York, NY, [4]Boston Children's Hosp., Boston, MA, [5]Ocean State Pediatrics, East Greenwich, RI, [6]Dept. of Pediatrics, Warren Alpert Med. Sch. of Brown Univ., Hasbro Children's/Rhode Island Hosp., Providence, RI, [7]Rhode Island and Hasbro Childrenâ□□s Hosp., Providence, RI, [8]Boston Children's Hosp. and Harvard Med. Sch., Boston, MA, [9]Manton Ctr. Gene Discovery Core, Boston, MA

Disclosure Block: **G. Iannello:** None.

Introduction: Lowe syndrome (LS) is a rare, genetic disorder caused by mutations in the X-linked gene *OCRL*. The cardinal clinical features are congenital cataracts, intellectual and developmental disability, and renal tubular dysfunction. *OCRL* codes for inositol polyphosphate 5-phosphatase, a protein that regulates several cellular processes including cell proliferation, cytoskeletal function, and membrane trafficking. *OCRL* has also been shown to localize in the primary cilium, and may be involved in cilia assembly through RhoGTPase signaling. We identified a boy with a frameshift mutation in *OCRL* without the renal and eye features of LS. This study aims to identify if the different phenotype could be explained by differential isoform expression across tissues. Materials and Methods: Whole exome sequencing of the trio was performed on Illumina platform with a mean coverage of ~100X (> 90% targets 20x). The Infinium Global Screening Array BeadChip 24v1.0 and WES data were used for calling copy number variants. The workflow in Seqr software was used for data analysis. To assess the transcript expression of *OCRL*, we used WT-fibroblast, renal cortex and neuroprecursor cells. RNA extraction was performed following the manufacturerâ□□s instructions, and qRT-PCR was conducted on Applied Biosystems QuantStudio Flex 7. Protein expression was assessed in the fibroblasts from the proband, WT control and LS patients by Western Blot (WB).

Results: We identified a novel hemizygous X-linked frameshift variant (c.2136_2139del+1insCT, p. L713Rfs*28) in *OCRL* in a 5-year old boy with extreme obesity (BMI 27 kg/m2, Z +5) and short stature (Z -2.2). The proband also had bilateral webbed toes, central hypoventilation, neurobehavioral dysfunction, and ganglioneuroblastoma. The variant segregates in the carrier mother, maternal aunt and maternal grandmother, and is absent in the unaffected father and maternal grand uncle. It is located in exon 18a (NM_000276), a 24bp region that is only present in the longer transcript of the *OCRL* gene (transcript a). Tissue expression profiling shows the presence of exon 18a in the neural precursor cells, but not in fibroblasts or renal cells. WB showed normal protein expression in the proband and WT, and absent in LS fibroblasts.

Conclusion: In this study, we report a new frameshift variant in *OCRL* in a proband with severe obesity, short stature and neurobehavioral phenotype, but no eye or renal disease. The differential isoform expression of OCRL in the brain and the peripheral tissues (fibroblasts and kidney) may explain the differences in phenotype. The unique location of the variant in this patient may help elucidate the role of *OCRL* in the brain.

PrgmNr 2792 - Adiponectin and cardiometabolic outcomes in sub-Saharan Africans: A Mendelian randomization study

View session detail

Author Block: **K. Meeks**, A. R. Bentley, A. Doumatey, A. A. Adeyemo, C. N. Rotimi; Ctr. for Res. on Genomics and Global Hlth., Natl. Human Genome Res. Inst., Bethesda, MD
Disclosure Block: **K. Meeks:** None.
Background: High adiponectin levels have been favorably associated with cardiometabolic outcomes in observational studies across populations. Mendelian randomization (MR) studies assessing the causality of these associations have generated contradictory results and MR studies in African ancestry populations are scarce. We performed MR analysis to assess the presence of a causal relationship between adiponectin and cardiometabolic outcomes in sub-Saharan Africans. In addition, we assessed whether this relationship differed for normal weight versus overweight/obese individuals.
Methods: We constructed a polygenic risk score (PRS) for circulating adiponectin levels using 3,393 unrelated Nigerian, Ghanaian, and Kenyan individuals from the Africa America Diabetes Mellitus (AADM) study. Summary statistics from the largest meta-analysis for genome-wide association studies on adiponectin to date were used as base data. The PRS was used as the instrumental variable in an MR analysis implemented using the two-stage least-squares method to assess its association with insulin resistance, high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), total cholesterol, triglycerides, systolic blood pressure, diastolic blood pressure, Type 2 Diabetes (T2D), and hypertension. Models were adjusted for age, sex, population stratification, Body Mass Index (BMI), alcohol intake, and smoking status. Analyses were conducted for the total sample and separately for normal weight (BMI 2) and overweight/obese (BMI â¥ 25 kg/m^2) individuals.
Results: Serum adiponectin levels were associated with all cardiometabolic outcomes. The constructed PRS contained 222 SNPs and had an F-statistic of 116.7, confirming that it is a strong instrumental variable. In MR analyses, the adiponectin PRS was associated with only LDL of the cardiometabolic traits considered (Î² = 25.2, 95%CI = 2.2 - 48.2, P = 0.032). The PRS was also associated with LDL in overweight individuals (Î² = 25.8, 95%CI = 0.1 - 51.4, P = 0.049). In normal weight individuals, the PRS was associated with T2D (OR 0.17, 95%CI = 0.03 - 0.96, P = 0.045). Epidemiological and causal estimates were directionally concordant.
Conclusions: The findings of this first MR study for adiponectin in sub-Saharan Africans support an influence on LDL in overweight individuals and T2D in normal weight individuals. Neither of the two previous MR studies that included other populations of African ancestry found evidence for a causal association with LDL, and one of the two found a causal role in T2D. Both genetic and environmental factors could be driving differences in association between populations.

PrgmNr 2793 - Analyzing human knockouts to validate *GPR151* as a therapeutic target for reduction of body mass index

View session detail

Author Block: **A. Gurtan**[1], J. Dominy[1], S. Khalid[2,3,4], L. Vong[1], S. Caplan[1], T. Currie[1], S. Richards[1], L. Lamarche[1], D. Denning[1], A. Rasheed[2], S. Hameed[5], Z-u. Asar[5], I. Saleem[5], A. Jalal[6], S. Abbas[6], S. Rasheed[7], F-u. Memon[8], M. Ishaq[7], P. Frossard[2], D. Saleheen[2,3,4]; [1]Novartis Inst.s for BioMed. Res., Cambridge, MA, [2]Ctr. for Non-Communicable Diseases, Karachi, Sindh, Pakistan, [3]Dept. of Med., Columbia Univ. Irving Med. Ctr., New York, NY, [4]Dept. of Cardiology, Columbia Univ. Irving Med. Ctr., New York, NY, [5]Punjab Inst. of Cardiology, Lahore, Pakistan, [6]Faisalabad Inst. of Cardiology, Faisalabad, Pakistan, [7]Karachi Inst. of Heart Diseases, Karachi, Pakistan, [8]Red Crescent Inst. of Cardiology, Hyderabad, Pakistan

Disclosure Block: **A. Gurtan:** Salary/Employment; Novartis.

Finding novel drug targets for sustained reduction in body mass index (BMI) may help curb obesity, which has become a global health risk that will affect half of the world population by 2030. Here, we show that GPR151 antagonism is not a compelling therapeutic approach for BMI reduction. To follow up on a previously published genetic association with BMI, we use the worldâ□□s largest biobank of human knockouts, the Pakistan Genome Resource (PGR), which has identified >14,000 knockouts for >5,000 genes through whole-exome sequencing. Previous studies from other cohorts have shown that heterozygous putative loss-of-function (pLoF) in *GPR151*, which encodes a G protein-coupled receptor, is associated with reduced BMI and reduced odds ratio of type 2 diabetes (T2D). To date, however, no homozygous pLoF carriers (human knockouts) have been characterized. Here, we test (i) if complete knockout of *GPR151* is compatible for survival in humans, and (ii) if *GPR151* pLoF is associated with a therapeutically meaningful (>5%) reduction in BMI. We investigate gene-dosage effect in *GPR151* pLoF heterozygotes and complete knockouts by analyzing BMI measurements (N = 27,329) and T2D status (6,561 cases, 33,162 controls). We identify 10 pLoF variants in *GPR151* with a cumulative allele frequency of 2.2%, comprising 38 knockouts and 1,141 heterozygous pLoF carriers. In vitro, GPR151 pLoF variant proteins are unstable and not expressed. In human knockouts, we do not identify any clinically meaningful reduction in BMI (26.3Â±4.4 kg/m^2 in knockouts versus 26.7Â±3.3 kg/m^2 in non-carriers). We observe a modest, nominally significant increase in T2D odds ratio in pLoF heterozygous carriers (odds ratio=1.2 [1.0 - 1.4], P=0.03). In a mouse model, *GPR151*$^{-/-}$ body weight is indistinguishable from wild-type on a standard chow diet. On a high fat diet, body weight of male *GPR151*$^{-/-}$ mice is elevated compared to wild-type. In total, complete loss of *GPR151* does not lead to BMI reduction in humans or in a preclinical mouse model.

PrgmNr 2794 - Circulating metabolite biomarkers for osteoporosis: Evidence from Mendelian randomization and direct metabolite measurement

View session detail

Author Block: **Y. Chen**[1,2], S. Zhou[1,2], V. Forgetta[3], A. Cerani[1], C. Langenberg[4,5], J. Richards[1,1,2]; [1]McGill Univ., Montreal, QC, Canada, [2]Lady Davis Inst., Jewish Gen. Hosp., Montreal, QC, Canada, [3]Lady Davis Inst., Montreal, QC, Canada, [4]Univ Cambridge, Cambridge, Cambridgeshire, United Kingdom, [5]CharitÃ© Univ. Med., Berlin, Germany
Disclosure Block: **Y. Chen:** None.
Objective: Circulating metabolites that causally influence bone density could be used as biomarkers or drug targets for osteoporosis. **Methods**: Here we undertook Mendelian Randomization (MR) studies to screen over 300 metabolites for their effects on estimated bone mineral density (eBMD) using data from three metabolite GWASs (up to N = 9,736) and one eBMD GWAS (N = 426,824). We then confirmed metabolites that were significantly associated with eBMD using sensitivity and colocalization analyses. Next, we tested the effects of these metabolites on DXA-measured BMD (up to N = 53,236) via additional MR studies. Last, we measured the prioritized metabolites and tested their association with hip total BMD using up to N = 7,109 plasma samples from the Canadian Longitudinal Study on Aging (CLSA) cohort. **Result**: MR results identified 40 metabolites estimated to influence eBMD (FDR 0.8) or had their genetic variants positionally close (within 1 Mb region) to a gene that encodes the enzyme or transporter for the metabolites. 4 of these metabolites (i.e., proline, pro-hydroxyproline, epiandrosterone sulfate, and 3-methyl catechol sulfate) were further replicated in the MR of DXA-measured BMD and the direct association test with total hip BMD in CLSA. Notably, using the different set of genetic variants selected from three metabolite GWASs, proline showed a consistent improving effect on eBMD and DXA-measured BMD. Evidence on the direct association of metabolites and hip total BMD using CLSA data also highlighted the effect of the proline pathway on bone density in humans.**Conclusion**: By integrating evidence from large-scale genetic data and direct metabolite measurements, we identified circulating metabolites and pathways that influence bone density in humans.

PrgmNr 2795 - Constructing Metabolic Syndrome-related polygenic scores in the multiethnic GENNID family study

View session detail

Author Block: **J. Y. Wan**, L. Simon, D. Robinson, A. R. Freedland, T. Norden-Krichmar, K. L. Edwards, American Diabetes GENNID Study Group; Dept. of Epidemiology and Biostatistics, Program in Publ. Hlth., Univ. of California, Irvine, CA
Disclosure Block: **J.Y. Wan:** None.
Objective: There is great interest in utilizing polygenic scores (PGS) to predict risk of common chronic conditions. PGS are often based on meta-analysis comprised of multiple genome-wide association analyses (GWAS) that emphasize common variants. It is not clear how well PGS based on GWAS from European ancestry (EA) populations will predict disease in non-EA populations. We use a multi-ethnic sample of families to: (1) evaluate the association between a common-variant genetic burden PGS and metabolic syndrome (MetS)-related traits; (2) to compare / contrast the prediction of MetS traits from PGS based under different scenarios.
Methods: The GENNID study consists of 1502 subjects in 259 families from European-American (EA), Mexican-American (MA), African-American (AA), and Japanese-American (JA) families. We used PLINK and R for data quality-control of the GENNID dataset, base GWAS summaries, and PGS analyses. Clumping and Thresholding (C + T) method was used to construct PGS based on single ancestry-specific GWAS meta-analysis summaries, summaries based on EA samples only, and from a meta-analysis consisting of multiple populations. We evaluated the association of PGS constructed under these different scenarios with BMI using a linear mixed effects model (with fixed covariate effects and random family effects). P-values were approximated using a normal distribution. For each ethnic group, ROC curves and AUC values (using bootstrapping methods) were used to compare the ability of PGS to predict obesity (BMI>30) and MetS traits.
Results: PGS based on a GWAS meta-analysis of multiple populations predicted better (i.e., had higher AUC) than PGS based on either ancestry-specific or EA GWAS. The AUC values ranged from 53% to 71% and explained less than 6% of the variance in obesity.
Conclusions: In general, the proportion of variance explained was low but is consistent with the literature for other common, complex conditions. Future analyses will evaluate whether including additional genomic variants identified through family-based methods will boost the predictive ability of a PGS.

PrgmNr 2796 - Enrichment of Motilin Receptor LOF variants in Gastroparesis

View session detail

Author Block: **J. Carlin**, S. Smieszek, M. Fisher, M. Bagger, D. Madonick, C. Kupersmith, M. Lopez-Lascurain, G. Birznieks, M. Polymeropoulos; Vanda Pharmaceuticals Inc., Washington, DC
Disclosure Block: **J. Carlin:** Salary/Employment; Vanda Pharmaceuticals Inc..
Gastroparesis is a serious chronic disorder characterized by delayed gastric emptying and upper gastrointestinal symptoms of nausea, vomiting, bloating, fullness after meals, abdominal discomfort and pain. To ascertain the genetic risk factors for gastroparesis, we conducted a large whole genome sequencing (WGS) study of idiopathic and diabetic gastroparesis subjects in a phase 3 study cohort. We previously reported an enrichment of Motilin Receptor (MLNR) gene LOF variants in WGS study of gastroparesis (phase 2 study VP-VLY-686-2301). Here we replicate this enrichment in an independent cohort of gastroparesis samples obtained from a phase 3 study VP-VLY-686-3301. We investigated the frequency and effect of rare loss-of-function (LOF) variants in adult subjects (n=320) with idiopathic or diabetic gastroparesis, evidence of delayed gastric emptying, and moderate to severe nausea. Patients were excluded with another active disorder or treatment which could contribute to gastroparesis symptoms including but not limited to gastric malignancy or neurological disorder or patients needing gastric or parental feeding.
In this dataset we report an increased frequency of a frameshift mutation within MLNR gene, variant rs562138828 in gastroparesis subjects compared to controls. Motilin is a 22 amino acid peptide hormone expressed throughout the gastrointestinal (GI) tract. The protein encoded by this gene is a motilin receptor which is a member of the G-protein coupled receptor 1 family. We have shown previously increased frequency of a frameshift mutation with MAF 0.01 as compared 0.0009 AF in GNOMAD (p-value=0.01). We reported 4/119 gastroparesis patients carry the variant of interest that results in p.Leu202ArgfsTer105. In the current data, the effect persists and we report 13 out of 894 versus controls 13 out of 1912 (MAF: cases 0.007; MAF: controls 0.003) and OR: 2.15. The variant has increased allelic frequency in African American population per GNOMAD controls, we inspected the frequencies in this group alone: MAF AA cases: 0.024 and MAF AA controls 0.01, OR: 2.36 (Relative risk: 1.71, p-value 0.02). The carriers were equally split among idiopathic and diabetic, so this possible gastroparesis risk factor appears to be agnostic as to diabetes status.
The identified LOF variants within the region can serve as a risk factor for disease as well as inform treatments, especially given accumulating evidence supporting different responses to treatment. The finding may be of direct relevance to individuals with the identified mutation as they may respond differently to gastroparesis treatments; especially those targeting MLNR.

PrgmNr 2797 - Exome-wide association analyses of MRI imaging-derived traits in up to 32k participants of the UK Biobank identify genes and variants involved in the regulation of body composition

View session detail

Author Block: **H. Kim**[1], J. Linge[2,3], O. Dahlqvist-Leinhard[2,3], J. Goodman[4], E. J. Keliher[4], C. L. Hyde[5], E. B. Fauman[1], M. R. Miller[1]; [1]Internal Med. Res. Unit, Pfizer, Cambridge, MA, [2]AMRA Med., LinkÃ¶ping, Sweden, [3]LinkÃ¶ping Univ., LinkÃ¶ping, Sweden, [4]Early Clinical Dev., Digital Med. and Imaging, Pfizer, Cambridge, MA, [5]Early Clinical Dev., Biostatistics, Pfizer, Groton, CT
Disclosure Block: **H. Kim:** Salary/Employment; Pfizer.
The combined analyses of imaging and genetic data offer an opportunity to study the genetic underpinnings of body composition traits such as adipose tissue volumes and ectopic fat deposits, which are important risk factors for cardiometabolic outcomes. In this study, we performed exome-wide association analyses (ExWAS) for 5 whole-body MRI imaging-derived traits in up to 32k individuals of European ancestry, combining currently available imaging data from 40k individuals and exome sequences from 450k individuals in the UK Biobank. From the MRI images, visceral adipose tissue (VAT), abdominal subcutaneous adipose tissue (ASAT), and thigh muscle volumes as well as liver and intramuscular fat percentages were quantified. Traits were adjusted for carefully selected sets of covariates and inverse normal transformed. We tested associations both at the individual variant level using BOLT-LMM and at the gene level, aggregating protein-altering variants into four variant sets and applying burden and SKAT tests. The analyses identified a total of 56 significant associations (P-value threshold adjusted for the number of variants or variant sets and for genomic inflation factor per trait) across 4 traits (no significant association for thigh muscle volume), including 47 associations of individual protein-altering (pLOF, missense, or inframe indel) variants and 9 associations of gene-level variant sets. All but one association overlapped with GWAS loci (defined by +/-500kb of significant variants) of the corresponding trait, including protein-altering variants in well-characterized genes such as *PNPLA3*, *TM6SF2*, *APOE*, *GCKR*, *GPAM*, *MTARC1*, *MTTP*, and *PPARG*. There was at least one significant ExWAS association in 28 GWAS loci, which can help inform potential causal genes for the GWAS signals. There was one association outside of GWAS loci: the burden association of the *APOB* pLOF variant set with liver fat (P-value = 6.1e-14, beta = 1.3 SD unit, aggregate MAF = 0.6%), which is consistent with the known function of *APOB* in VLDL secretion from the liver. Notably, in the *APOB* region, no individual variants from exome or imputed data had associations reaching exome- or genome-wide significance (minimum P-value in the region = 9.4e-5). This study shows that ExWAS of imaging-derived traits has the potential to identify new genes and variants not captured by GWAS and help prioritize causal genes and variants in the GWAS loci. We expect that the statistical power of ExWAS for imaging traits will increase further as the number of individuals with both imaging and exome data reaches the projected goal of 100k individuals.

PrgmNr 2798 - G6PD deficiency associated genetic variants and the diagnostic utility of HbA1c: the Africans in America Study

View session detail

Author Block: **K. Ekoru**[1], A. A. Adeyemo[1], A. R. Bentley[1], C. W. DuBose[2], A. C. Patterson[2], A. Wentzel[2], C. N. Rotimi[1], A. E. Sumner[2,3]; [1]Ctr. for Res. on Genomics and Global Hlth., Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [2]Section on Ethnicity and Hlth., Natl. Inst. of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, [3]Natl. Inst. of Minority Hlth.and Hlth.Disparities, Bethesda, MD

Disclosure Block: **K. Ekoru:** None.

Glucose-6-phosphate dehydrogenase (G6PD) deficiency is an X-linked red cell enzymopathy associated with increased risk of hemolysis. Genetic variants in G6PD have been reported to be associated with non-glycemic lowering of HbA1c with potential impact on diagnosis and monitoring of diabetes and related phenotypes. Within the context of the Africans in America Study, we evaluated the diagnostic utility of glycated hemoglobin A1c (HbA1c) for detecting abnormal glucose tolerance (AGT) [2-hour postprandial glucose â¥7.8 mmol/L] in 350 African-born blacks living in the United States [65% male, mean age 40 (SD 10) years, mean body mass index (BMI) 27.9 (SD 4.4) kg/m2] with and without the G6PD A- (202A/376G) deficiency haplotype. (African region of origin: West 50%, Central 17%, East 33%). Forty percent [40% (141/350)] of the participants had AGT [8% diabetes and 32% prediabetes] by the oral glucose tolerance test (OGTT), and 16% (56/350) carried the G6PD A- (202A/376G) haplotype. In logistic regression models adjusted for age, sex and BMI, the 202A/376G haplotype was strongly associated with AGT as defined by HbA1c [OR = 0.045 (95% CI 0.010-0.211), p= 8.056 X 10^{-5}]. On the other hand, a similar model for AGT defined by the OGTT showed a much weaker though still significant association [OR = 0.483 (95% CI 0.248-0.942), p=0.033]. Among individuals carrying the 202A/376G haplotype, the model-based Receiver Operating Characteristic (ROC) curve for the ability of HbA1c to detect AGT-OGTT was below the reference line, indicating a diagnostic accuracy worse than random classification. In contrast, the ROC curve was above the reference line among individuals carrying the wild-type genotype [AUC = 0.62 (95% CI 0.55 - 0.69)]. These findings show that the G6PD A- (202A/376G) deficiency haplotype is strongly associated with AGT, with a bigger effect on AGT defined by HbA1c versus AGT defined by OGTT. Notably, HbA1c has poor utility for diagnosing both prediabetes and diabetes in the presence of the G6PD A- (202A/376G) deficiency haplotype. These findings add to the list of factors that affect the utility of HbA1c for the diagnosis and monitoring of glycemic status and highlights the need to develop additional glycemic biomarkers or combine HbA1c with other markers such as glycated albumin or fasting glucose.

PrgmNr 2799 - Genome-wide assessments of obesity and 138 inflammatory markers to characterize obesity-induced inflammation

View session detail

Author Block: **J. Li**[1], J. Hu[2], X. Han[1], J. E. Manson[3], F. B. Hu[1], L. Liang[4]; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [4]Harvard Sch Publ. Hlth., Boston, MA
Disclosure Block: **J. Li:** None.
Chronic inflammation induced in obesity plays an etiological role in the development of obesity-related diseases such as coronary heart disease (CHD), stroke, type 2 diabetes (T2D), and colorectal cancer (CRC). **However**, specific inflammatory molecular/cellular profiles induced by obesity and their genetic basis are largely unknown, hindering the identification of therapeutic targets that ameliorate obesity-related diseases. **We aimed to** identify biomarkers for obesity-induced inflammation and investigate the shared genetic architecture between obesity and such an inflammatory profile. **Pooling multiple databases**, we curated the largest genome-wide association summary statistics for 138 markers including 126 inflammatory proteins (C-reactive protein [CRP] N>450k, other proteins N~4k-45k) and 12 blood immune cell types (N>450k) and subsets (N~2.5k); and for body mass index (BMI) and waist-hip ratio (WHR, with/without adjusting for BMI; N~695k-807k), CHD, stroke, T2D, and CRC (N~522k-898k). **In Mendelian randomization analyses**, both BMI and WHR showed significant causal effects for 6 inflammatory markers (*FDR*PFDRFDRGenome-wide cross-trait meta-analyses identified 2329 independent genetic loci significantly coregulating obesity (BMI and/or WHR) and at least one marker of obesity-induced inflammation (e.g., *ACAD10* coregulated BMI and 8 markers, *P*enriched expressions in blood vessels and brain tissues, while loci coregulating WHR and WHR-induced inflammation harbored genes highly expressed in adipose tissue, blood, and blood vessels (*FDR*Finally, we identified markers of obesity-induced inflammation showing significant causal effects for obesity-related diseases (e.g., IL6 for CHD, stroke, and T2D; *FDR*In conclusion, we identified a comprehensive profile of molecular/cellular markers for obesity-induced inflammation and characterized genetic architecture between obesity and such a profile. Our data indicate shared and distinct inflammatory pathways induced by obesity and central obesity, and suggest specific pathways contributing to obesity-related diseases.

PrgmNr 2800 - Identifying novel high-impact rare disease-causing mutations, genes and pathways in exomes of Ashkenazi Jewish inflammatory bowel disease patients

View session detail

Author Block: **Y. Itan**[1], Y. Wu[1], K. Gettler[2], M. Giri[1], K. Sabic[1], L. Schumm[3], M. Daly[4], R. H. Duerr[5], M. S. Silverberg[6], J. D. Rioux[7], S. R. Brant[8], D. P. McGovern[9], J. H. Cho[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Mt. Sinai Sch. of Med., Fort Myers, FL, [3]Univ. of Chicago, Chicago, IL, [4]Harvard Med. Sch., [5]Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA, [6]Mount Sinai Hosp, Toronto, ON, Canada, [7]Montreal Heart Inst. | Res. Ctr., Montreal, QC, Canada, [8]Rutgers Robert Wood Johnson Med. Sch., New Brunswick, NJ, [9]Los Angeles, CA

Disclosure Block: **Y. Itan:** None.

Inflammatory bowel disease (IBD) is a group of chronic digestive tract inflammatory conditions whose genetic etiology is still poorly understood. The incidence of IBD is particularly high among the Ashkenazi Jews (AJ). Here, we identified 7 novel and plausible IBD-causing genes from the exomes of 4,974 genetically identified AJ IBD cases (1,905) and controls (3,069). Various biological pathway analyses were then performed, along with bulk and single-cell RNA sequencing, to demonstrate the likely physiological relatedness of the novel genes to IBD. Importantly, we demonstrated that the rare and high impact genetic architecture of AJ adult IBD displays a significant overlap with very early onset IBD genetics. Moreover, by performing phenome-wide analyses in major biobanks, we found that IBD genes have pleotropic effects that involve other immune responses. Finally, we showed that polygenic risk score analyses based on genome-wide high impact variants have high power to predict IBD susceptibility.

PrgmNr 2801 - Integrating genetics and plasma proteomics to predict risk of type 2 diabetes in adults and children: A two-sample Mendelian randomization study

View session detail

Author Block: **F. Ghanbari**[1], N. Yazdanpanah[1], M. Yazdanpanah[1], D. Manousaki[2]; [1]Univ. de Montreal, montreal, QC, Canada, [2]Sainte Justine Hosp., Montreal, QC, Canada
Disclosure Block: **F. Ghanbari:** None.
Type 2 diabetes (T2D) is a global pandemic in adults, and its prevalence has increased substantially in children in the past two decades, despite stabilization of the obesity rates. In both adults and children, primary prevention of T2D through lifestyle modification is key, and identification of biomarkers enabling early disease prediction is important for selecting candidates for these interventions. Proteins are key factors that transfer information from genome to phenome, can be measured in the blood using high-throughput proteomic platforms, and represent promising drug targets. Therefore, identification of circulating protein biomarkers that are causal in T2D in adults and children can be used to enhance prediction and understanding of the mechanism underlying this disease. Here, we applied a large scale two-sample Mendelian randomization (MR) study, using *cis* genetic determinants of n=1,089 circulating proteins for adult T2D and n=174 proteins for youth-onset T2D from five large genome-wide association studies (GWAS). We screened these proteins for causal associations with adult T2D risk in 74,124 T2D cases and 824,006 controls of European ancestry, and with youth-onset T2D risk in 664 T2D cases and 1,976 controls of European ancestry from the largest available GWAS on these traits. Our MR analysis identified 20 candidate proteins for adult T2D, among which 6 proteins (MANS domain containing 4 [MANSC4], Histo-blood group ABO system transferase [ABO], Haptoglobin [HP], Sodium/potassium-transporting ATPase subunit beta-2 [ATP1B2], Spermatogenesis-associated protein 20 [SPATA20], C-type mannose receptor 2 [MRC2]) shared a common causal variant with adult T2D in colocalization analyses. Moreover, we identified 11 proteins with evidence of causal association with youth-onset T2D. In summary, using an unbiased MR approach, our study leveraged large-scale proteomic and T2D GWAS data to identify circulating proteins as candidate biomarkers and drug targets for adult and youth-onset T2D.

PrgmNr 2802 - Integrative multiomics analyses reveals link between obesity and insulin resistance-associated plasma metabolites and adipose tissue transcripts in African Americans

View session detail

Author Block: **S. K. Das**[1], N. D. Palmer[1], M. E. Comeau[1], N. K. Sharma[1], C. Zhang[2], A. Mardinoglu[2], D. W. Bowden[1], L. E. Wagenknecht[1], M. C. Ng[3], C. D. Langefeld[1]; [1]Wake Forest Sch. of Med., Winston-Salem, NC, [2]KTH-Royal Inst. of Technology, Stockholm, Sweden, [3]Vanderbilt Univ. Sch. of Med., Nashville, TN

Disclosure Block: **S.K. Das:** None.

Metabolic changes initiated by obesity and insulin resistance (IR) are risk factors for type 2 diabetes (T2D), a disease diagnosed by increased plasma glucose levels. Analyses of plasma metabolite levels has facilitated the understanding of biochemical mechanisms of obesity and IR, but such studies in non-European ancestry cohorts remain limited. Derangement of gene expression in tissues involved in glucose homeostasis are likely to be mechanistically linked with altered plasma metabolite levels in obesity and IR, but remain understudied.To address these issues, we integrated measures of insulin sensitivity (Matsuda index from OGTT, and S_I from FSIGT) and obesity (BMI), fasting plasma metabolite profiles (liquid chromatography-mass spectrometry, Metabolon), and adipose and muscle tissue gene expression (Illumina HT12-v4) data in 253 non-diabetic African Americans (AA) from AAGMEx cohort.Among the 1124 plasma metabolites evaluated in AAGMEx participants, 166, 153 and 60 were significantly associated (FDR-PI, respectively. Lipid metabolites 1-oleoyl-GPC and cortolone glucuronide and amino acids hypotaurine and glycine were among the top ranked BMI and IR-associated metabolites in AAGMEx. Analyses were similarly undertaken in AAs from the IRASFS cohort (N~565) which replicated directionally consistent significant association of 100 and 36 metabolites for BMI and S_I, respectively. Glycine was positively associated, while palmitoyl-linoleoyl-glycerol (a Diacylglycerol) was inversely associated with S_I in both cohorts. Hierarchical clustering of BMI and IR-associated metabolites defined the clusters of correlated metabolites. GWAS of metabolites selected from these clusters, identified SNPs in genetic loci putatively regulating plasma levels of a subset of these metabolites in AAGMEx. Reporter metabolite analyses by integrating genome-scale metabolic model for human, and adipose tissue transcript-trait association statistics in AAGMEx predicted significant correlations of 60, 48, 51 metabolites with BMI, Matsuda index and S_I, respectively. Empirically measured plasma metabolite data in AAGMEx validated 12 metabolites. Plasma mannose levels were inversely correlated with Matsuda index (b= -1.4, FDR-P = 7.37E-7), and in adipose, expression of genes involved in mannose metabolism (*GMPPA*, *GMPPB*, and *GMDS*) were also inversely correlated with Matsuda index. In summary, this study delineates the architecture of obesity and IR-associated metabolites in African Americans, and also suggests a role for altered adipose tissue gene-expression in determining at least a subset of these metabolites.

PrgmNr 2803 - Machine Learning for Metabolite Estimation to Examine Contributors to Adiposity: The GUARDIAN Consortium

View session detail

Author Block: **N. D. Palmer**[1], H. C. Ainsworth[2], X. Guo[3], A. Mackay[4], F. Nyasimi[5], D. W. Bowden[6], Y. Liang[7], K. D. Taylor[8], L. J. Raffel[9], T. A. Buchanan[4], R. M. Watanabe[4], J. I. Rotter[10], L. E. Wagenknecht[11], C. D. Langefeld[1], H. Im[5]; [1]Wake Forest Sch. of Med., Winston-Salem, NC, [2]Wake Forest Sch. of Med., Winston-salem, NC, [3]The Lundquist Inst., Torrance, CA, [4]Keck Sch. of Med. of USC, Los Angeles, CA, [5]The Univ. of Chicago, Chicago, IL, [6]Wake Forest Univ Sch Med, Winston-Salem, NC, [7]Univ. of Chicago, Chicago, IL, [8]The Lundquist Inst. for BioMed. Innovation, 1124 W. Carson Street, Torrance, CA 90502, Torrance, CA, [9]Los Angeles, CA, [10]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [11]Wake Forest Univ Sch Med., Winston-Salem, NC

Disclosure Block: **N.D. Palmer:** None.

Obesity is a complex disease characterized by extensive metabolic dysregulation. Metabolomic profiles are highly informative capturing the interaction of cellular processes and environmental exposures to promote disease. Therefore, metabolomics has the potential to improve mechanistic understanding and identify clinically relevant targets. However, use of this technology can be limited by cost and sample availability. To facilitate metabolite investigations into complex disease, data from the Insulin Resistance Atherosclerosis Family Study (IRASFS, n=943 Mexican Americans) with empiric metabolomics data (Metabolon HD4) and genome-wide genotype data were used. 950 metabolites were significantly heritable (P2 up to 0.87). Ridge and LASSO regression revealed cross validated correlation between predicted and observed measures >0.1 for 448 metabolites. In an independent set of 192 Mexican Americans, the LASSO regression (enforcing sparsity by selecting a small proportion of variants) yielded better estimation performance than the ridge penalty (full polygenic architecture) suggesting that the genetic architecture of metabolites in this sample was sparse. Most metabolites had multiple independent loci (1-22, median=21) contributing to the estimation. Estimation was extended to the six Mexican American cohorts in the GUARDIAN Consortium to assess their association with measures of adiposity. Multiple significant (FDR P2=0.47) and was negatively associated with BMI, WC and WHR (FDR P=0.014-8.5E-6). In addition, three uncharacterized metabolites also exhibited high heritably (h^2>0.56) and significant association with measures of adiposity (FDR P=0.014-3.1E-6) suggesting this approach could contribute to metabolite characterization. In conclusion, we have developed genetic algorithms for estimating 448 metabolites with good performance in independent datasets and found strong evidence of a sparse genetic architecture. This estimation model was used to assess association with adiposity phenotypes highlighting known and novel associations, further demonstrating utility. Future work will expand genetic coverage to include rare variants and phenotypic associations to further characterize metabolic dysregulation.

PrgmNr 2804 - Mapping genomic regulation of kidney diseases and traits at a cell type and variant level of specificity

View session detail

Author Block: **C. J. Benway**, S. Han, M. T. McNulty, A. Greenberg, D. Lee, Nephrotic Syndrome Study Network (NEPTUNE), M. Sampson; Boston Children's Hosp., Boston, MA
Disclosure Block: **C.J. Benway:** None.
Although numerous genetically associated loci for kidney function and disease have been identified by genome-wide association studies (GWAS), determining the causal genes and functional variants remains a major challenge. Since the majority of trait-associated variants map to intergenic and non-coding genomic regions, integration of GWAS results with other data types (such as expression quantitative trait loci [eQTL]) can help identify causal and functional variants in a tissue- or cell-type specific manner. Further, analysis of disease tissue (rather than healthy tissue) may uncover context-specific associations that may otherwise not be detectable. To more precisely resolve the regulatory landscape of kidney diseases and traits, explain heritability, and identify potential mechanisms, we integrated eQTL data from micro-dissected glomerular (n = 240) and tubulointerstitial (n = 311) RNA-seq transcriptomes from individuals with nephrotic syndrome (NS) and summary statistics from two large trans-ethnic GWAS meta-analyses for estimated glomerular filtration rate (eGFR) and urinary albumin-to-creatinine ratio (UACR).
Utilizing a Bayesian statistical framework for eQTL discovery and multi-SNP fine-mapping (TORUS/DAP), we identified 5,526 glomerular eQTLs and 9,742 tubulointerstitial eQTLs at 50%) for eGFR. This result includes replication of gene expression associations for *UMOD* and *FGF5*, as well as novel associations to *LARP4B* and *RRAGD* which can be attributed to colocalization at a single variant for each locus. We identified 7 tubulointerstitial and 16 glomerular co-localization signals (RCP > 50%) for UACR. In addition to replicating co-localization signals at *PRKCI* and *TGFB1* in glomerular tissue, we refined the co-localization signal at *PTH1R* to a single variant, rs6787229, which also co-localized with expression of *MYL3*.

PrgmNr 2805 - Metabolomic signatures of waist circumference in the Hispanic Community Health Study/ Study of Latinos (HCHS/SOL)

View session detail

Author Block: **V. Buchanan**[1], R. Rohde[1], H. M. Highland[1], B. Yu[2], C. R. Isasi[3], C. L. Avery[1], R. Kaplan[3], J. I. Rotter[4], M. L. Daviglus[5], L. Van Horn[6], J. Cai[7], Q. Qi[3], E. Boerwinkle[8], R. J. F. Loos[9], S. G. Buyske[10], K. Ferrier[11], K. E. North[1], K. L. Young[1]; [1]Univ. of North Carolina at Chapel Hill, Dept. of Epidemiology, Chapel Hill, NC, [2]Univ. of Texas Hlth.Sci. Ctr. at Houston, Dept. of Epidemiology, Human Genetics, and Environmental Sci., Houston, TX, [3]Albert Einstein Coll. of Med., Dept. of Epidemiology and Population Hlth., Bronx, NY, [4]Inst. for Translational Genomics and Population Sci., The Lundquist Inst. and Dept. of Pediatrics, Harbor-UCLA Med. Ctr., Torrance, CA, [5]Univ. of Illinois Coll. of Med., Inst. for Minority Hlth.Res., Chicago, IL, [6]Northwestern Univ., Feinberg Sch. of Med., Chicago, IL, [7]Univ. of North Carolina at Chapel Hill, Dept. of Biostatistics, Chapel Hill, NC, [8]Univ. of Texas Hlth.Sci. Ctr. at Houston, Dept. of Epidemiology, Human Genetics and Environmental Sci., Houston, TX, [9]The Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [10]Rutgers Univ., Dept. of Statistics, Piscataway, NJ, [11]Univ. of Colorado Denver, Aurora, CO
Disclosure Block: **V. Buchanan:** None.

Elevated central adiposity is a risk factor for cardiometabolic disease (CMD), and its prevalence has increased in the US over the past three decades, with stark disparities among minorities. While the genetic impact on central adiposity is well-established, metabolomic dysregulation in the pathogenesis of central obesity, which may point to dysregulated genomic pathways that underlie disease manifestation, remains understudied. We thus aim to identify metabolites associated with waist circumference (WC) and downstream CMD in a large ancestrally diverse cohort of Hispanic Americans.

HCHS/SOL is a community-based prospective cohort of Hispanic/Latino adults from randomly selected households at 4 US field centers. Interviews and clinical assessment were conducted at in-person visits from 2008-2011 (Visit 1). A total of 3972 participants were randomly selected for metabolomic profiling (using Visit 1 fasting blood draws) using an untargeted platform. We investigated cross-sectional associations between metabolites and WC adjusted for BMI, age, study center, background group, and sample weights from Visit 1 using linear regression models, first stratified by sex and then meta-analyzed.

Data for 687 known metabolites and WC were available for 3806 participants (mean age: 45.8 years, 57% female). We identified Bonferroni-corrected significant associations (pThis work will help in better understanding the pathways leading from genes to central obesity in a diverse, highly admixed population. Identification of central obesity-related metabolites may provide novel insights concerning the metabolic and physiologic mechanisms underlying central obesity.

PrgmNr 2806 - Methylation profiles at birth linked to early childhood obesity

View session detail

Author Block: **S. J. Craig**[1], D. Lariviere[1], I. M. Paul[2], M. E. Marini[1], F. Chiaromonte[1], K. D. Makova[1], M. L. Reimherr[1]; [1]Penn State Univ., University Park, PA, [2]Penn State Coll. of Med., Hershey, PA
Disclosure Block: **S.J. Craig:** None.
Obesity affects 18.5% of children and adolescents in the United States. Previous studies have suggested an overlap of the underlying mechanisms of obesity across the lifecourse. It has also been shown that obesity is linked to a widespread change in the methylation profile in peripheral blood. Based on this, we aimed to measure how changes in methylation profile in mothers and newborns at birth are correlated to the childâ□□s growth during the first three years after birth and to the probability of a child to be overweight. We used the data from a subset of the SIBSight (Effects of Birth Order and Genetics on Infant Parenting and Obesity Risk) birth cohort that followed approximately 117 second born children from birth to age 3 years. In addition to genetic profiles, the study tracked the childâ□□s growth and environmental factors such as nutrition quality and familial environment. The analysis presented here focused on 48 children whose methylation profiles were measured at birth using Illumina Infinium MethylationEpic chips. The methylation profiles were evaluated by sampling placenta tissue and cord blood. We performed regression analyses to identify genes with differential methylation patterns associated with children's growth from birth to age 2 years. We show that the two types of tissue studied have different methylation patterns and highlight different sets of genes linked to childrenâ□□s growth index. The majority of the genes identified here were linked to weight-related traits in previous studies through GWAS and variant analyses. For instance, they were linked to body mass index (BMI), hip-to-waist ratio, or nutrient assimilation. Our results suggest a link between the methylation profile during pregnancy and the prevalence of obesity in early childhood.

PrgmNr 2807 - Multi-ethnic genome-wide meta-analysis in type 1 diabetes

View session detail

Author Block: **D. A. Michalek**[1], C. C. Robertson[1], W-M. Chen[1,2], B. M. Delelegn[1], T1DGC, S. Onengut-Gumuscu[1,2], S. S. Rich[1,2]; [1]Ctr. for Publ. Hlth.Genomics, Univ. of Virginia, Charlottesville, VA, [2]Dept. of Publ. Hlth.Sci., Univ. of Virginia, Charlottesville, VA
Disclosure Block: **D.A. Michalek:** None.
Background: Type 1 diabetes (T1D) is an autoimmune disorder caused by destruction of pancreatic Î²-cells. The original GWAS meta-analysis (2009) and fine mapping have identified over 50 loci associated with risk of T1D, yet limited GWAS have been completed over the last decade. **Objective:** Discover new T1D loci utilizing a multi-ethnic GWAS approach. **Methods:** A total of 3,222 families (11,476 individuals, majority of European ancestry), 891 African ancestry (AFR, 409 T1D cases, 482 controls), and 308 Admixed individuals (AMR, 153 T1D cases, 155 controls) were genotyped on the Illumina CoreExome BeadChip. A total of 430,930 SNPs passed quality control metrics with additional SNPs and insertion/deletion (indel) polymorphisms obtained from imputation (TOPMed). Family data was analyzed with the Generalized Disequilibrium Test (GDT), and case-control data employed logistic regression (PLINK). Meta-analysis (METAL) identified SNPs and indels associated with T1D with P -8. **Results:** Following imputation, 8,711,967 variants (MAF > 0.01) were used in meta-analysis. Eight loci were associated with T1D, including the MHC (rs9273364, P = 2.94 x 10^{-501} in *HLA-DQB1*), *INS* (rs3842753, P = 1.35 x 10^{-49}, in complete LD with known *INS* rs689), *PTPN22* (rs2476601, P = 3.45 x 10^{-26}), *SH2B3* (rs3184504, P = 1.05 x 10^{-18}), *ERBB3* (rs705708, P = 1.60 x 10^{-14}), *RNLS* (rs3781197, P = 4.01 x 10^{-12}) and *CTLA4* (rs3087243, P = 1.82 x 10^{-9}). A new locus, *NRP1* (rs2666237, P = 3.61 x 10^{-8}) on 10p11.22 was identified. Two regions (*FAM43A*, 3q29, novel; *IL2RA*, 10p15.1, known) had suggestive (P -7) evidence of association with T1D. Neuropilin-1, encoded by the *NRP1* gene, is a marker for murine CD4+FoxP3+ regulatory T (Treg) cells. *NRP1* also positively regulates T cell proliferation. *FAM43A* is expressed in NK cells, naÃ¯ve B cells, monocytes and in certain subgroups of CD4+ T cells (DICE project) yet little is known about its function and relationship with T1D. **Conclusions:** These findings suggest that inclusion of subjects with diverse genetic ancestry can identify new and refine known regions of the genome contributing to risk of T1D, and the results can aid in understanding the biology underlying type 1 diabetes.

PrgmNr 2808 - Novel Genetic Loci Associated with Age-of-onset of Type 1 Diabetes

View session detail

Author Block: **H-Q. Qu**[1], J. Qu[1], J. P. Bradfield[2], X. Chang[3], L. Marchand[4], J. Glessner[5], Y. Liu[6], M. March[7], J. Li[1], J. J. Connolly[1], J. D. Roizen[8], P. M. Sleiman[3], C. Polychronakos[9], H. Hakonarson[7]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Quantinuum Res. LLC, San Diego, CA, [3]CHOP, Philadelphia, PA, [4]McGill Univ., Montreal, QC, Canada, [5]Children s Hosp. of Philadelphia, Philadelphia, PA, [6]The children's Hosp. of philadelphia CHOP, Philadelphia, PA, [7]Children S Hosp. of Philadelphia, Philadelphia, PA, [8]The Children S Hosp. of Philadelphia, Philadelphia, PA, [9]McGill Univ. Hlth.Ctr., Montreal, QC, Canada

Disclosure Block: **H. Qu:** None.

Type 1 diabetes (T1D) has been traditionally recognized as an autoimmune disease, and numerous autoimmune genes have been identified for underlying the genetic susceptibility of T1D. A minor proportion (~5-10%) of Caucasian patients diagnosed as T1D have non-autoimmune pathogenesis, i.e. T1bD(Leslie, Atkinson et al. 1999). For most patients with T1D, the pancreatic destruction is irreversible once T1D is diagnosed. Prediction of T1D onset and early intervention have been warranted for research efforts. Sharp et al. developed a specific genetic risk scoring (GRS) system for T1D, (T1D-GRS2), using 67 single nucleotide polymorphisms (SNPs) from known autoimmune loci(Sharp, Rich et al. 2019), which demonstrated excellent performance for the prediction of T1D genetic susceptibility. As shown in the Type 1 Diabetes Genetics Consortium (T1DGC) GWAS cohort, the GRS2 scores are also correlated with the T1D age-of-onset (r=-0.102, *P*= 8.35E-11). By a large collection of European T1D cases and controls, we defined high GRS2 vs low GRS2 at GRS2=8.43 with the sensitivity of 0.855 and the specificity of 0.719 for T1D prediction. In further, our genetic association study showed distinct genetic architecture between cases with high GRS2 and those with low GRS2, suggesting the phenotypic heterogeneity of T1D (Qu, Hui-Qi et al. 2021). Based on this GRS definition, 3510 (86.9%) of 4037 T1DGC cases had high GRS, and 527 (13.1%) had low GRS. GRS2 scores in both GRS groups are correlated with the T1D age-of-onset (high GRS2 group r= -0.064, *P*=1.54E-04; low GRS2 group r= -0.093, *P*=0.033). In our study based on the imputation by the TOPMed (Version R2 on GRC38) Reference Panel, we searched for genetic association with T1D age-of-onset in the two GRS groups separately, corrected for GRS2 scores and population structure. In the high GRS group, four independent association signals with T1D age-of-onset were identified at the *HLA DR/DQ* region, i.e. rs602457 (*P*=1.98E-13), rs9273371 (*P*=5.42E-09), rs4713573 (*P*=5.58E-09), and rs35353422 (*P*=8.76E-09); and one independent signal at the *HLA* class I region was associated with T1D age-of-onset, i.e. rs116020851 (*P*=1.93E-08). In contrast, in the low GRS group, the only one association signal with genome-wide significance was not from the *HLA* region, but from the *SLC22A31* gene region at Chr16, tagged by rs4785604 (*P*=2.83E-08). *SLC22A31* has biased expression in lung and pancreas. The SNP rs4785604 is associated with the expression level of *SLC22A31* in brain, testis, and spleen (https://www.gtexportal.org). In addition to the potential of T1D onset prediction, these novel loci highlight the heterogeneous pathogenesis of high GRS T1D and low GRS T1D.

PrgmNr 2809 - Pleiotropic Loci Influencing Cystic Fibrosis-Related Diabetes and Neonoatal Intestinal Obstruction

View session detail

Author Block: **M. Atalar Aksit**[1], H. Ling[2], R. G. Pace[3], A. V. Faino[4], K. S. Raraigh[1], F. Onchiri[4], A. Stilp[5], E. E. Blue[5], Y-H. Zhou[6], E. W. Pugh[7], M. J. Bamshad[5], R. L. Gibson[5], M. R. Knowles[3], G. R. Cutting[1], S. M. Blackman[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]1812 Ashland Ave, Suite 200, Baltimore, MD, [3]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [4]Seattle Childrenâ□□s Res. Inst., Seattle, WA, [5]Univ. of Washington, Seattle, WA, [6]North Carolina State Univ., Raleigh, NC, [7]Johns Hopkins Univ, Baltimore, MD

Disclosure Block: **M. Atalar Aksit:** None.

Introduction: Cystic fibrosis (CF), a life-limiting autosomal recessive disorder, has a range of phenotypes that vary considerably amongst individuals, including CF-related diabetes (CFRD) and neonatal intestinal obstruction (meconium ileus; MI). Though MI and CFRD are not correlated, several loci are â□□pleiotropicâ□□, i.e., associate with both traits. Here, we investigate the degree of genetic overlap between MI and CFRD using data from the CF Genome Project (CFGP). **Methods:** Whole genome sequencing was performed on enrollees with CF from 5 studies (CF Twin and Sibling Study, CFRD Study, Genetic Modifier Study [GMS], GMS of Severe CF Liver Disease Study, and Early Pseudomonas Infection Control Observational Study). 4,473 individuals with exocrine pancreatic insufficient CF (905 with MI; 1,410 with CFRD) were studied. MI was modeled as a binary variable via mixed-effects logistic regression, adjusted for site, birth cohort, residual CFTR function (zero vs. nonzero), 4 principal components, and genetic relatedness. CFRD was modeled as a mixed-effects linear regression for Martingale residuals for age of onset, with the same covariates as MI plus sex, as it associated with CFRD. Variants with MAF>0.5% were analyzed. **Results:** Common variants at *SLC26A9*, *CEBPB* and *PRSS1* were significantly associated with both MI and CFRD: the direction of effect was shared at *SLC26A9* and *CEBPB* (the risk allele increases risk of CFRD and MI), but was discordant at *PRSS1*. The *PRSS1* variants, which increase risk for MI and are protective for CFRD, are also protective for pancreatitis in non-CF individuals (Derikx, et al. 2015.). High-risk variants for *CEBPB* also increase risk for type 2 diabetes in non-CF individuals (Klupa, et al. 2000). A less common (MAF 1%) missense variant in *SLC26A9 (V172M)* associated with MI and not CFRD. Variants associated with MI but not CFRD at *CLPS/SLC26A8*, *EMSY* and *SLC6A14* while variants at *TCF7L2* and 16p12.1 associated with CFRD but not MI. **Conclusions:** Both MI and CFRD are influenced by multiple genetic variants, some of which associate with only one of the two traits, some which influence both traits in the same direction, and one of which associates in opposite directions. *Supported by CFF grants CUTTIN18XX1, BAMSHA18XX0, and KNOWLE18XX0 and submitted on behalf of the CF Genome Project.*

PrgmNr 2810 - Polygenic susceptibility to obesity modifies the impact of salivary amylase gene (*AMY1*) copy number on BMI

View session detail

Author Block: **L. Stalbow**[1,2,3], R. A. J. Smit[1,2,3,4], M. H. Preuss[1,2,3], A. Moscati[1], N. Patel[5,3], P. Garg[5,3], A. J. Sharp[5,3], R. J. F. Loos[1,2,3,6,7]; [1]The Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]The Genetics of Obesity and Related Metabolic Traits Program, Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]The Mindich Child Hlth.and Dev. Inst., Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Dept. of Clinical Epidemiology, Leiden Univ. Med. Ctr., Leiden, Netherlands, [5]The Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]The Dept. of Environmental Med. and Publ. Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, [7]NNF Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, Denmark
Disclosure Block: **L. Stalbow:** None.
Background: As GWASs get larger and are able to explain more of the genetic variation in body mass index (BMI), polygenic risk scores (PRS) can better identify individuals with a genetic susceptibility for high BMI. These scores are based on GWAS summary statistics and cannot capture more complex genetic architecture, such as copy number (CN) variations, and are therefore missing some of the genetic structure of BMI. For example, low CN of the salivary amylase gene (*AMY1*) has been associated with an increase in BMI in many, but not all studies. This inconsistency may be ascribed to the overall polygenic environment, influencing the impact of *AMY1* CN on BMI. Here we examine how high BMI polygenic susceptibility exacerbates the effect of *AMY1* CN on BMI.
Methods: Through exome sequencing of 6,653 individuals of European ancestry (EA) of the Mount Sinai Bio*Me* biobank, we determined *AMY1* CN using CNVnator on a subset of Bio*Me* with whole genome sequencing, and extrapolating the results to the exome sequencing files. We generated a BMI PRS using recent GWAS summary statistics from ~700,000 EA participants. We estimated the association between *AMY1* CN and the PRS (both standardized) with BMI using linear regression analyses. We then divided the PRS into a high group (top third of the PRS) and low group (lower two thirds of the PRS) and examined the interaction between *AMY1* CN and the PRS groups. Analyses were adjusted for age, sex and the first ten principal components.
Results: A 1-standard deviation (SD) increase in PRS was associated with a 1.4 kg/m^2 (SE 0.07, P AMY1 CN was not significantly associated with BMI (-0.087 kg/m^2 (SE 0.065, P=0.18)). There was some evidence that the PRS affects the association between *AMY1* CN and BMI ($P_{interaction}$ = 0.21), which was most pronounced at the highest PRS-level ($P_{interaction}$ = 0.002), i.e., the effect of *AMY1* CN on BMI was different between individuals in the high vs low PRS groups. Specifically, in the high PRS group, a 1-SD increase in *AMY1* CN was significantly associated with a -0.37 kg/m^2 (SE 0.13, P=0.005) decrease in BMI (or 1.1 kg in body weight for a 1.7m tall person), whereas in the low PRS group, a 1-SD increase *AMY1* CN was not associated with BMI (0.06 kg/m^2 (SE 0.074, P=0.4)), and the direction of effect was reversed.
Conclusion: We show that in the presence of a high polygenic susceptibility for increased BMI, low AMY1 CN contributes to an increase in BMI. Interestingly, this effect was not observed in individuals with a low PRS. These findings contribute to the discussion as to the effect of AMY1 CN on BMI, and highlights the interplay between CN and polygenic susceptibility.

PrgmNr 2811 - Role of diet and submaximal cardiorespiratory fitness as mediators of genetic susceptibility to obesity: results from the Quebec Family Study

View session detail

Author Block: **L. PÃ©russe**[1], R. Jacob[1], C. Bertrand[1], C. H. Llewellyn[2], A. Tremblay[1], C. Bouchard[3], M-Ã. LabontÃ©[1], V. Drapeau[1]; [1]Laval Univ., Quebec, QC, Canada, [2]Univ. Coll. London, LONDON, United Kingdom, [3]Pennington BioMed. Res Ctr, Baton Rouge, LA
Disclosure Block: **L. PÃ©russe:** None.
Background : Genome-wide association studies (GWAS) have identified a large number of genetic variants associated with body mass index (BMI). The mechanisms by which these genetic variants contribute to the increased risk of obesity can be investigated using mediation analysis, an approach used to assess the extent to which the effect of an exposure (e.g. genetic susceptibility) on an outcome (e.g. obesity) is explained by a given set of hypothesized mediators (also called intermediate variables). **Objective:** This study aimed to assess the role of diet and submaximal cardiorespiratory fitness (CRF) as potential mediators of the relationship between genetic susceptibility to obesity (assessed using a polygenic risk score or PRS) and BMI and waist circumference (WC). **Methods:** This cross-sectional study included 750 adult participants from the Quebec Family Study. A PRS of obesity based on >500,000 BMI genetic variants was calculated using the LDpred2 software. Dietary intakes were assessed with a 3-day food record from which intakes of various food groups and a diet quality score (i.e., Nutrient Rich Food Index 6.3) were derived. CRF was determined by an incremental submaximal exercise test performed on ergocycle. The mediating effect ($Î²_{indirect}$) of diet and CRF on the association between the PRS of obesity and BMI and WC was assessed using a regression-based and bootstrapping approach after adjustment for age, sex, menopause, smoking and dieting and misreporting (for dietary data). **Results :** The PRS explained 25.7% and 19.8% of the variance in BMI and WC, respectively. The association between PRS and BMI was partly mediated by a low CRF ($Î²$=0.74Â±0.21, p=0.0008), high intakes of sodium ($Î²$=0.74Â±0.22, p=0.0002), fat and high-fat foods ($Î²$=0.46Â±0.16, p=0.002) and sugar-sweetened beverages ($Î²$=0.25Â±0.14, p=0.03), low intake of fruits ($Î²$=0.37Â±0.12, p=0.006) and poor diet quality ($Î²$=0.33Â±0.12, p=0.009). Similar results were observed for WC. **Conclusion:** The results show that genetic susceptibility to obesity is partly mediated by a poor diet quality, undesirable intake of specific food groups and a low CRF, suggesting that improvements in diet quality and exercise capacity may reduce obesity risk among individuals with high genetic susceptibility.

PrgmNr 2812 - Therapeutic validation of *FATP5*/*SLC27A5* using loss-of-function variation in humans

View session detail

Author Block: **S. Khalid**[1,2], P. Bhatnagar[3], T. P. Ryan[3], A. Rasheed[2], Z. Asar[4], I. Saleem[4], A. Jalal[5], S. Abbas[5], S. Z. Rasheed[6], S. Hameed[4], F-u. R. Memon[7], P. M. Frossard[2], M. Ishaq[6], D. Saleheen[1,2]; [1]Dept. of Med., Columbia Univ. Irving Med. Ctr., New York, NY, [2]Ctr. for Non-Communicable Diseases, Karachi, Pakistan, [3]Eli Lilly & Company, Indianapolis, IN, [4]Punjab Inst. of Cardiology, Lahore, Pakistan, [5]Faisalabad Inst. of Cardiology, Faisalabad, Pakistan, [6]Karachi Inst. of Heart Diseases, Karachi, Pakistan, [7]Red Crescent Inst. of Cardiology, Hyderabad, Pakistan

Disclosure Block: **S. Khalid:** None.

FATP5 (SLC27A5) is a fatty acid transporter, expressed almost exclusively in the liver, which facilitates uptake of long-chain fatty acids in hepatic cells and is required for recycling of bile acids. Deletion of FATP5 in mice has shown to lower triglyceride levels in the liver and protect against obesity, through a reduction in food intake. Importantly, knocking out FATP5 in mice with diet induced NASH, has also shown to reverse the NASH phenotype and improve glucose homeostasis. Given the increasing disease burden of NASH and obesity worldwide, pharmacological inhibition of FATP5 could be an attractive therapeutic target. Here, we used human genetics approach utilizing Pakistan Genetic Resource (PGR) and other publicly available biobanks (UK Biobank and Biobank Japan) to assess therapeutic validity of FATP5, and associated benefits/risks for metabolic syndromes. We meta-analyzed FATP5 common variants for hepatic, cardiovascular, glycemic, and inflammatory serum biomarkers (n=21) along with anthropometric traits (BMI and waist circumference). Total 15 biomarkers/traits were observed associated with an FDR

PrgmNr 2813 - Trans-ancestry meta-analyses for waist-hip ratio in >1.1M individuals

View session detail

Author Block: **E. Wilson**[1], S. Vedantam[2,3], E. Marouli[4], J. D. Arias[5], G. Chittoor[6], S. I. Berndt[5], T. W. Winkler[7], K. L. Young[8], M. Graff[8], C. T. Liu[9], C. M. Lindgren[10,11], K. L. Mohlke[1], A. E. Justice[6], Genetic Investigation of ANthropometric Traits (GIANT) Consortium; [1]Dept. of Genetics, Univ. of North Carolina, Chapel Hill, NC, [2]Dept. of Endocrinology, Boston Children's Hosp., Boston, MA, [3]Broad Inst. of MIT and Harvard, Cambridge, MA, [4]William Harvey Res. Inst., Barts and The London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom, [5]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD, [6]Dept. of Population Hlth.Sci., Geisinger Hlth.System, Danville, PA, [7]Dept. of Genetic Epidemiology, Univ. of Regensburg, Regensburg, Germany, [8]Dept. of Epidemiology, Univ. of North Carolina, Chapel Hill, NC, [9]Dept. of Biostatistics, Boston Univ., Boston, MA, [10]Wellcome Trust Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom, [11]Oxford Big Data Inst., Li Ka Shing Ctr. for Hlth.Information and Discovery, Univ. of Oxford, Oxford, United Kingdom
Disclosure Block: **E. Wilson:** None.

GWAS meta-analyses of individuals of diverse ancestry incorporating allelic effect heterogeneity across populations can identify new loci. We compared results from fixed-effect and mixed-effect models used to perform trans-ancestry meta-analyses for waist-to-hip ratio adjusted for body mass index (WHRadjBMI). Our combined sample size was >1.1 million individuals, including 947K, 136K, 41K, 33K, and 28K individuals of European, East Asian, South Asian, Hispanic/Latino, and African populations, respectively, from 201 studies that contributed GWAS results to the GIANT consortium. We performed a fixed-effects trans-ancestry meta-analysis including 92M variants (5M indels) using RAREMETAL and a mixed-effects trans-ancestry meta-analysis using MR-MEGA with three axes of variation as covariates for allelic effects to better account for allelic heterogeneity across populations. In addition, to evaluate potential inflation due to population stratification, we calculated lambda-GC values with three MAF thresholds (>0%, 0.5%, and 5%) and estimated linkage disequilibrium score regression (LDSR) intercepts and attenuation ratios using ~3M common variants. For comparison between methods, we identified distinct loci (P -9) based on distance (+/- 500 kb) from index/lead variants. Also, we compared these counts to a previous WHRadjBMI GWAS meta-analysis of 694K individuals of European ancestry that identified 346 loci (Pulit, 2019). The lambda-GC values from RAREMETAL and MR-MEGA ranged between 1.09-1.26 and 1.07-1.29, respectively. In general, lambda-GC values were lower for MR-MEGA than RAREMETAL, consistent with MR-MEGAâ□□s ability to account for heterogeneity between populations. However, based on LDSR among the ~80% European-ancestry individuals, only ~1.4% of the inflation in test statistics could be attributed to stratification, suggesting that GC values are inflated due to polygenicity. Trans-ancestry analyses using RAREMETAL without correction identified 611 loci, while using MR-MEGA identified 500 loci. Compared to the 118 loci identified only by RAREMETAL (8 index variants with MAF = 0.01) showed greater allelic heterogeneity across populations (mean chi-square het 20.8+/- 5.5 vs 2.4+/-1.8 for RAREMETAL). While the MR-MEGA analysis identified fewer loci overall, it detected some loci missed under an assumption of homogeneous effects. Using either model, our trans-ancestry meta-analysis had a 71% larger sample size than the previous European-focused WHRadjBMI meta-analysis and identified over 150 new loci.

PrgmNr 2814 - *IL10RB* as a key regulator of COVID-19 host susceptibility and severity

View session detail

Author Block: **G. Voloudakis**[1,2], G. E. Hoffman[1], S. Venkatesh[3,2], K. M. Lee[4], K. Dobrindt[1], J. M. Vicari[1], W. Zhang[1], N. D. Beckmann[5], S. Jiang[1], D. Hoagland[1], J. Bian[4], L. Gao[6], A. Corvelo[7], K. Cho[8], J. S. Lee[9], S. K. Iyengar[10], S-W. Luoh[11], S. Akbarian[1], R. Striker[12], T. L. Assimes[13], E. E. Schadt[1], M. Merad[1], B. R. tenOever[1], A. W. Charney[1], Mount Sinai COVID-19 Biobank, VA Million Veteran Program COVID-19 Science Initiative, K. Brennand[1], J. A. Lynch[14], J. F. Fullard[1], P. Roussos[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]James J. Peters VA Med. Ctr., Bronx, NY, [3]Icahn Sch. of Med. at Mount Sinai, Armonk, NY, [4]VA Salt Lake City Hlth.Care System, Salt Lake City, UT, [5]New York, NY, [6]Oregon Hlth.& Sci. Univ., Portland, OR, [7]New York Genome Ctr., New York, NY, [8]VA Boston Hlth.care System, Boston, MA, [9]Stanford Univ. Sch. of Med., Stanford, CA, [10]Case Western Reserve Univ, Cleveland, OH, [11]lake oswego, OR, [12]Univ. of Wisconsin, Madison, WI, [13]Stanford Univ Sch Med., Stanford, CA, [14]Salt Lake City VA Med. Ctr., Bedford, MA
Disclosure Block: **G. Voloudakis:** None.
Background/Aims: Recent efforts have identified genetic loci that are associated with coronavirus disease 2019 (COVID-19) infection rates and disease outcome severity. Translating these genetic findings into druggable genes and readily available compounds that reduce COVID-19 host susceptibility is a critical next step. **Methods:** We integrate COVID-19 genetic susceptibility variants, multi-tissue genetically regulated gene expression (GReX) and perturbargen signatures to identify candidate genes and compounds that reverse the predicted gene expression dysregulation associated with COVID-19 susceptibility. The top candidate gene is validated by testing both its GReX and observed blood transcriptome association with COVID-19 severity, as well as by in vitro perturbation to quantify effects on viral load and molecular pathway dysregulation. We validate the *in silico* drug repositioning analysis by examining whether the top candidate compounds decrease COVID-19 incidence based on epidemiological evidence. **Results:** We identify *IL10RB* as the top key regulator of COVID-19 host susceptibility. Predicted GReX up-regulation of *IL10RB* and higher *IL10RB* expression in COVID-19 patient blood is associated with worse COVID-19 outcomes. In vitro *IL10RB* overexpression is associated with increased viral load and activation of immune-related molecular pathways. Azathioprine and retinol are prioritized as candidate compounds to reduce the likelihood of testing positive for COVID-19. **Conclusions:** We establish an integrative data-driven approach for gene target prioritization. We identify and validate *IL10RB* as a suitable molecular target for modulation of COVID-19 host susceptibility. Finally, we provide evidence for a few readily available medications that would warrant further investigation as drug repositioning candidates.

PrgmNr 2815 - A genome-wide association study identifies two novel loci for respiratory infection with *Pseudomonas aeruginosa* in cystic fibrosis

View session detail

Author Block: **B. Lin**[1], J. Gong[2], N. Panjwani[3], K. Keenan[3], C. Wang[4], L. Sun[1], L. J. Strug[5]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]the Hosp. for Sick Children, Toronto, ON, Canada, [3]The Hosp. for Sick Children, Toronto, ON, Canada, [4]Toronto, Canada, [5]Hosp for Sick Children, Toronto, ON, Canada
Disclosure Block: **B. Lin:** None.
Pseudomonas aeruginosa (*Pa*) is a common pathogen that contributes to progressive Cystic Fibrosis (CF) lung disease. Genetic factors contribute approximately 50% to 85% of the variation in age of chronic *Pa* infection in CF individuals but remain unknown. We conducted a genome-wide association study of genetic modifiers for the age of first and chronic *Pa* infection in Canadian individuals with CF. Our primary analysis on 1,037 individuals identified two novel genome-wide significant loci, *rs62369766* (near *FGF10* on Chromosome 5; *P* value=1.78E-8) and *rs927553* (*SPATA13* on Chromosome 13; *P* value=1.72E-8), for chronic *Pa* infection age. Through a phenome-wide association study of *rs62369766* and *rs927553* in population-based databases, we observed their association with lung function and immunological phenotypes, respectively in non-CF cohorts. We further investigated the genetic overlap between chronic *Pa* infection age and lung function in CF through a polygenic risk score (PRS), defined in the largest GWAS of CF lung disease to date (n=6,365). CF lung function has a moderate phenotypic correlation with chronic *Pa* infection age (Pearson correlation coefficient=0.12, *P* value=2E-4). The PRS constructed from ~8,000 SNPs associated with CF lung function is significantly associated with chronic *Pa* infections age (*P* value = 0.006), supporting the presumption that targeting some genetic factors associated with lung function will delay the onset of chronic infections. Our study identifies novel loci potentially modifying the age of chronic *Pa* infections in CF, and provides new insights into the genetic basis of *Pseudomonas aeruginosa* infections.

PrgmNr 2816 - Aberrant levels of CXCL16 in severe COVID19 patients

View session detail

Author Block: **S. Smieszek**; Vanda Pharmaceuticals, DC, DC
Disclosure Block: **S. Smieszek:** None.
Genome-wide association studies have recently identified 3p21.31, encompassing the *CXCR6* gene, as one of the strongest thus far reported susceptibility loci for hospitalized COVID-19 (rs73064425, OR=2.14, discovery p=4.77 Ã 10^{-30} Pairo-Castineira et al., 2020). Nakanishi et al., reported that risk allele carriers experienced an increased risk of COVID-19 related mortality/ complications: severe respiratory failure (odds ratio [OR] 2.0, 95%CI 1.6-2.6), venous thromboembolism (OR 1.7, 95% CI 1.2-2.4). CXCL16 is synthesized as a transmembrane molecule that is expressed as a cell surface-bound molecule, and as a soluble chemokine. CXCL16 interacts with CXCR6 promoting chemotaxis or cell adhesion. CXCL16 has been previously implicated in the pathogenesis of lung injury and functions as a chemoattractant for CXCR6+ T cells. The CXCR6/CXCL16 axis mediates homing of T cells to the lungs in disease and hyper-expression is associated with localized cellular injury. To characterize the CXCR6/CXCL16 axis in the pathogenesis of severe COVID-19, plasma concentrations of CXCL16 from 114 hospitalized COVID-19 patients participating in ODYSSEY COVID-19 clinical trial (and 37 controls) were assessed. CXCL16 levels in plasma were determined with ELISA kit (Themofisher). We also obtained samples for WGS analysis as well as for viral genome sequencing. We report a significant difference in plasma CXCL16 between COVID-19 hospitalized cases and controls (p-valueCXCR6 and *CXCL16* was observed in severe COVID-19 compared to mild disease and significant functional polymorphisms in *CXCR6* were linked to viral control. Our current study further supports crucial role of the CXCR6/CXCL16 axis in the immunopathogenesis of severe COVID-19.

PrgmNr 2817 - Adding polygenic risk scores drastically improves the diagnostic accuracy of conventional lab tests for systemic lupus erythematosus

View session detail

Author Block: **C. Khunsriraksakul**[1], Q. Li[2], M. Patrick[3], R. Sauteraud[4], D. J. McGuire[5], B. Li[6], L. C. Tsoi[7], D. J. Liu[8]; [1]Penn State Coll. of Med., Hershey, PA, [2]Univ. of Michigan, Ann Arbor, MI, [3]Univ. of Michigan, Ann Arbor, [4]Penn State Univ., Hershey, PA, [5]Penn State Coll. of Med., Hummelstown, PA, [6]Vanderbilt Univ, Nashville, TN, [7]Univ MICHIGAN, Ann Arbor, MI, [8]Penn State Univ., hershey, PA
Disclosure Block: **C. Khunsriraksakul:** None.
Systemic lupus erythematosus (SLE) is a systemic, remitting, and relapsing autoimmune disease that predominantly affects young women of childbearing age. The treatment option for SLE is rather limited. The diagnosis of SLE is heavily relied on the clinical classification criteria, and traditional lab tests used for SLE lack either sensitivity or specificity. It remains challenging for accurate diagnosis. Importantly, a late diagnosis of SLE is often associated with a worse prognosis (and even fatality), and therefore, accurate early diagnosis and effective intervention are critical for the patients. Here, we explore the benefits of incorporating polygenic risk score (PRS) in conjunction with other clinical features for SLE diagnosis. To create PRS models, we directly utilized the most recent summary statistic from SLE genome-wide association study (GWAS) of 4,943 cases and 8,483 controls (Julia et al., 2018) and further created another joint summary statistic by applying multi-trait analysis of GWAS (MTAG) to summary statistics of SLE and a few other related syndromes (e.g. rheumatoid arthritis, systemic sclerosis). We, then, derived a total of twelve candidate PRS models using six different PRS methods, including SBLUP, LDpred-inf, LDpred-funct, LASSOSUM, PRS-CS-auto, and SBayesR. We tested the validity of each PRS model in Vanderbilt University Biobank (BioVU; N = 49,707 and SLE prevalence = 0.35%). PRS-CS-auto with MTAG joint summary statistics achieve the best area under the curve (0.79) for the receiver operating characteristic curve. We found that individuals with PRS above the 98th percentile confer at least >5x increased risk for SLE compared to the general population. We further investigated the added benefits of PRS, when used in conjunction with routine clinical lab tests for SLE diagnosis (e.g. anti-nuclear antibody (ANA) and anti-dsDNA). Our result indicated that PRS can improve the stratification of patients with certain classical SLE lab results. For individuals that are ANA-positive or anti-dsDNA-negative, PRS further distinguished cases from controls, and thus improved the accuracy of conventional lab tests. Specifically, ANA-positive and anti-dsDNA-negative individuals with PRS above 95th percentile and 90th percentile were 20 and 12 times more likely to have SLE than that with low PRS. Finally, we validated our PRS models in Michigan Genomics Initiative and the results were consistent across all scenarios. We anticipate that genomics-informed precision medicine will facilitate physicians to give a more accurate, confident, and earlier diagnosis of this convoluted and hard-to-diagnose disease lupus.

PrgmNr 2818 - Application of a novel instrument-free and microfluidics-free single-cell analysis technology (PIPseq) that is well suited for viral applications in resource constrained laboratories

View session detail

Author Block: **R. Meltzer**[1], J. Turcinovic[2], J. Connor[2], A. May-Zhang[1], A. Osman[1], S. Kiani[1], K. Fontanez[1]; [1]Fluent BioSci., Watertown, MA, [2]Boston Univ., Boston, MA
Disclosure Block: **R. Meltzer:** Salary/Employment; Fluent BioSciences.
Single cell RNA sequencing (scRNA-seq) has made profound impacts in the study of cellular and molecular diversity in complex tissues. In the study of virology, this resolution of single-cell transcriptional changes in response to viral infection provides valuable insight into the mechanisms of infection and host response. Current scRNA-seq methods are not easily adopted in the virology lab as they are expensive, require complex instrumentation and consumables, and hence can be challenging to implement in a laboratory with limited resources and accessibility. Fluent BioSciences has developed a novel scRNA-seq approach with Pre-templated Instant partitions (PIPseq) that enables the analysis of thousands of cells without requiring complex instrumentation and consumables. The small format and convenient workflow, with the lack of instrumentation, allows PIPseq to be easily implemented in high-containment laboratories. Using a GFP-expressing control virus, we have demonstrated simultaneous capture and barcoding of viral and cellular transcriptomes, identified cellular gene expression shifts in response to viral infection, and identified gene expression responses in non-infected cells from low MOI infected samples compared to mock controls. Furthermore, we have demonstrated that the PIPseq protocol is effective at inactivating residual virus in sequencing library samples, thus enabling convenient sample post-processing outside of the high containment virology laboratory. Overall, we demonstrate that PIPseq is an effective method to study viral-host interactions at a single-cell resolution. The ongoing COVID-19 pandemic exemplifies the need for new tools and methods to elucidate the mechanisms of viral infection, pathogen-host responses, and diversity in cellular responses to infection.

PrgmNr 2819 - Assessing the potential correlation of polymorphisms in the *IL6R* gene with relative IL-6 elevation in severely ill COVID-19 patients

View session detail

Author Block: **A. Kaden**, J. L. Brzezynski, J. A. Shinn, M. M. Gibson, C. Xiao, S. P. Smieszek, C. Polymeropoulos, V. Polymeropoulos, G. Birznieks, M. H. Polymeropoulos; Vanda Pharmaceuticals Inc., Washington, DC
Disclosure Block: **A. Kaden:** Salary/Employment; Vanda Pharmaceuticals Inc..

Elevated levels of interleukin-6 (IL-6) may play an important role in the pathophysiology of COVID-19, yet attenuated responses are not seen across all severe patients. We aimed to determine the effect of baseline IL-6 levels, as well as other clinical variables, on outcomes (including mortality) in hospitalized COVID-19 patients. Additionally, we aimed to explore the genetic variants associated with attenuated IL-6 responses. Blood samples for baseline IL-6 levels and whole genome sequencing (WGS) analyses were collected from hospitalized patients enrolled in ongoing Vanda COVID-19 clinical studies.

We report significantly elevated levels of IL-6 in COVID-19 infected hospitalized patients (n = 100; p In prior literature, several variants including IL6-174 G/C (rs1800795 A), a variant in the promoter of the *IL-6* gene, have been reported to affect IL-6 serum levels. Interestingly, in our study the minor allele seemed to correlate with mortality (n = 150). We also inspected WGS data from our COVID-19 patients for the presence of loss-of-function *IL-6* variants, and we did not report any in this study cohort. Additionally, we tested the association between variants in the *IL-6* IL6R region and IL-6 plasma levels. Linear regression corrected for principal components and covariates identified a strong significant signal driven by the known exonic variant rs2228145 C, previously shown to affect IL-6 serum levels; the minor allele of this variant is associated with higher serum IL-6 levels.

While it is unlikely that a â□□cytokine stormâ□□, where the immune system produces an overabundance of cytokines, is seen in the majority of patients with severe COVID-19, baseline elevations of IL-6 greater than 150 pg/mL may be associated with the worst clinical outcomes and may warrant specialized treatment considerations. So far, no clinical studies have used IL-6 baseline assessments to stratify patient populations. We believe that careful examination and interpretation of IL-6 levels and genetic variants can help determine the patient population with the most likely robust clinical response to IL-6 inhibition.

PrgmNr 2820 - Comparative genetic architecture of the inflammatory bowel disease across East Asian and European populations

View session detail

Author Block: **R. Liu**[1,2], H. Gao[3], M. Li[4], H. Huang[2,1], Z. Liu[3]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [3]Ctr. for IBD Res., Dept. of Gastroenterology, Shanghai Tenth Peopleâ□□s Hosp. of Tongji Univ., Shanghai, China, [4]Dept. of Gastroenterology, Third Affiliated Hosp. of Guangzhou Med. Univ., Guangzhou, China

Disclosure Block: **R. Liu:** None.

The inflammatory bowel diseases (IBD) are a group of chronic autoimmune disorders including two major subtypes: Crohnâ□□s disease and ulcerative colitis. The prevalence and incidence of IBD are increasing worldwide, especially in East Asia and other developing countries. Genome-wide association studies have identified over 240 of risk loci for IBD. However, the majority of IBD genetic studies were conducted using subjects of European descent (EUR), limiting the discovery and application of IBD genetics to the rest of the world populations. To address this issue, we conducted a large-scale IBD genetic study in the East Asian populations (EAS) using subjects from China, Japan and Korea, for a total sample size close to 18,000 (CD: 4,323, UC: 3,648, and Control: 10,014). All subjects were genotyped on the Illumina Immunochip or the Asia Screening Array, and have undergone stringent quality controls and imputation. Using this cohort, we found 33 genetic loci associated with IBD beyond genome-wide significance, among which 17 had never been reported in IBD genetic studies in EAS. Two of these new IBD-associated loci in EAS are implicated by coding variants in *RUNX3* (P=2.9e-8) and *ADAP1* (P=2.8e-8). *RUNX3* is a transcription factor playing an important role during the development of T cells and regulating TGFÎ² signaling. Runx3 knockout mice spontaneously develop IBD. *ADAP1* is involved in the BCR Signaling Pathway. We also found a pleiotropic variant (P=1.1e-8), located in the intron of *GTF21*, associated with both IBD and systemic lupus erythematosus. Across the genome, we found common variants underlying IBD genetic risk have similar genetic effects between EAS and EUR ancestries, with a genetic correlation of 0.95. We also found the odds ratios of IBD putative causal variants highly consistent across ancestries (slope: 0.9). Despite the overall consistency and in line with previous reports, we found a few loci showing clearly different genetic effects in EAS vs EUR, including *TNFSF15* and *IL23R*, suggesting gene-environment interactions modifying the genetic risks across populations. Encouraged by the overall consistency of genetic effects across populations, we performed a fixed-effect meta-analysis with the latest European IBD GWAS and identified over 50 new loci associated with IBD. In summary, we have demonstrated the value of including diverse ancestries in IBD genetics research through building and leveraging the largest IBD genetics cohort of non-European ancestry. Through joint and comparative analyses with European IBD GWAS, we have identified over 50 new IBD genetic loci and revealed important insights into the genetic IBD epidemiology across populations.

PrgmNr 2821 - COVID-19 Dynamic pQTLs: Genetic Determinants of Proteins that are Unique to SARS-CoV-2 Infection

View session detail

Author Block: **S. Zhou**[1], T. Nakanishi[1], E. Brunet-Ratnasingham[2], G. Butler-Laporte[3], D. Morrison[1], L. Laurent[1], V. Forgetta[1], V. Mooser[4], D. E. Kaufmann[2], B. Richards[1]; [1]Lady Davis Inst., Jewish Gen. Hosp., McGill Univ., Montreal, QC, Canada, [2]Res. Ctr. of the Ctr. Hosp.ier de lâ□□UniversitÃ© de MontrÃ©al, Montreal, QC, Canada, [3]Lady Davis Inst., Jewish Gen. Hosp., McGill Univ., MontrÃ©al, QC, Canada, [4]McGill Genome Ctr., Montreal, QC, Canada

Disclosure Block: **S. Zhou:** None.

SARS-CoV-2 leads to profound changes in protein abundances. These changes vary widely between samples, suggesting that there are drivers of protein levels that are activated during infection, and these changes may be influenced by the host genome. Identifying these critical control points for changes in levels of these protein targets could lead to new targets for drug development. We therefore performed proteomic GWASs of 4,701 protein targets measured on the SomaLogic platform in two cohorts: 190 COVID-19 European ancestry patients during acute infection stage; and 160 European ancestry individuals who were post-COVID-19 infection or were SARS-CoV-2 PCR negative. We identified pQTLs (p-11) associated with levels during acute COVID-19 infection for 718 proteins, of which 66 were cis-pQTLs. All 66 cis-pQTLs were identified in healthy cohorts. A total of 590 proteins with trans-pQTLs were found only in COVID-19 patients during infection, while absent from proteomic GWAS of SARS-CoV-2 PCR negative individuals or large population cohorts. Among these acute infection de-novo trans-pQTLs, 45 proteins showed differential expression in patients between acute infection and post-infection phases. Some of these proteins, such as IRF9, PLK1 and TMPRSS11A have important roles in immune response. In conclusion, we identified a set of proteins whose genetic determinants might change during acute SARS-CoV-2 infection. Such proteins are likely important control points for the host response to COVID-19.

PrgmNr 2822 - Data-driven analysis exploring age and sex effects on gene expression for dengue hemorrhagic fever

View session detail

Author Block: **L. Rogers**; Univ. of the Virgin Islands, St Thomas, Virgin Islands, U.S.
Disclosure Block: **L. Rogers:** None.
Background: Dengue fever (DF) is a mosquito-borne disease caused by the dengue virus (DENV). DENV is prevalent in tropical/subtropical regions and has four known serotypes. It is also the most pervasive infection transmitted by *Aedes* mosquitoes. Dengue hemorrhagic fever (DHF) is a fatal complication of DENV infection, with hemorrhaging and vascular leakage symptoms. In addition, DHF often occurs because of secondary DENV infections. Our study analyzes pre-existing DF expression data and defines and explores how expression profiles may be affected by sex, age, and DF disease severity.
Methods: We are analyzing pre-existing protein array data from Brazil (12 subjects: 6 severe acute DF; 6 mild acute DF) and corresponding expression data after DF recovery. We use a linear mixed-effects model to investigate multiple study factors, identify differentially expressed genes, and assess disease severity, age, and sex effects on DF. We are also conducting a meta-analysis to determine expression profile differences between dengue-infected and healthy subjects. We are using publicly available gene expression datasets that we are curating from Array Express, Immune Space, and Gene Expression Omnibus (GEO). Our analysis focuses on datasets where the donor's disease state, age, and gender are reported. We are pre-processing the data by performing background correction, data normalization, and batch effect correction. We will also use a linear model to identify differentially expressed genes and explore the effects of our study factors (age, sex, disease state) on DF gene expression profiles.
Results: Our approach will highlight significant gene signatures and pathways associated with DF. Furthermore, we will identify potential gene targets for improving current treatments and exploring genes expressed equally across ages and sex when considering vaccinations for DENV.
Conclusion: Finally, analyzing host systemic responses can help characterize and differentiate between gene expression profiles for non-severe and severe DF.

PrgmNr 2823 - Downregulation of exhausted cytotoxic T cells in gene expression networks of multisystem inflammatory syndrome in children

View session detail

Author Block: **N. Beckmann**[1], P. Comella[1], E. Cheng[1], L. Lepow[1], A. G. Beckmann[1], S. Tyler[1], M. Konstantinos[1], N. Simons[1], G. E. Hoffman[1], N. Francoeur[1], on behalf of the Mount Sinai COVID19 biobank, S. Gnjatic[1], B. D. Gelb[2], M. Merad[1], R. Sebra[1], E. E. Schadt[1], A. Charney[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Mindich Child Hlth.and Dev. Inst., New York, NY
Disclosure Block: **N. Beckmann:** None.
Multisystem inflammatory syndrome in children (MIS-C) presents with fever, inflammation and pathology of multiple organs in individuals under 21 years of age in the weeks following severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. Although an autoimmune pathogenesis has been proposed, the genes, pathways and cell types causal to this new disease remain unknown. Here we perform RNA sequencing of blood from patients with MIS-C and controls to find disease-associated genes clustered in a co-expression module annotated to $CD56^{dim}CD57^{+}$ natural killer (NK) cells and exhausted $CD8^{+}$ T cells. A similar transcriptome signature is replicated in an independent cohort of Kawasaki disease (KD), the related condition after which MIS-C was initially named. Probing a probabilistic causal network previously constructed from over 1,000 blood transcriptomes both validates the structure of this module and reveals nine key regulators, including TBX21, a central coordinator of exhausted $CD8^{+}$ T cell differentiation. Together, this unbiased, transcriptome-wide survey implicates downregulation of NK cells and cytotoxic T cell exhaustion in the pathogenesis of MIS-C.

PrgmNr 2824 - Fine-mapping of multiple sclerosis susceptibility loci in a multi-ethnic population

View session detail

Author Block: **A. H. Beecham**[1], L. Gomez[1], S. Caillier[2], C. P. Manrique[1], P. Calabresi[3], K. Fitzgerald[3], N. A. Patsopoulos[4], D. Woo[5], S. Delgado[6], A. Chinea[7], L. Amezcua[8], J. Oksenberg[2], J. L. McCauley[1]; [1]John P. Hussman Inst. for Human Genomics, Univ. of Miami, Miami, FL, [2]Dept. of Neurology, Univ. of California at San Francisco, San Francisco, CA, [3]Dept. of Neurology, Johns Hopkins Univ., Baltimore, MD, [4]Brigham & Women's Hosp, Boston, MA, [5]Dept. of Neurology and Rehabilitation Med., Univ. of Cincinnati Coll. of Med., Cincinnati, OH, [6]Multiple Sclerosis Div., Dept. of Neurology, Univ. of Miami, Miami, FL, [7]San Juan MS Ctr., Guaynabo, PR, [8]Dept. of Neurology, Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA

Disclosure Block: **A.H. Beecham:** None.

Although 200 autosomal genetic variants outside of the Major Histocompatibility Complex (MHC) have been identified for association with multiple sclerosis (MS) risk; minimal work has been done to fine-map surrounding regions and pinpoint causal variants which may have a direct effect on disease pathology. While the discovery effort which identified the 200 MS risk variants included only populations of European ancestry, recent studies have shown substantial replication of risk variants in Hispanics and African Americans. Fine-mapping in genetically diverse populations is advantageous given the lower levels and distinct patterns of linkage disequilibrium (LD). Our goal was to fine-map MS risk loci in populations of African Americans (1676 MS cases, 1511 controls) and Hispanics (2399 MS cases, 2622 controls) who were genotyped on a genome-wide array (Illumina Infinium Multi-Ethnic Global + ImmunoArray-24 v2 BeadChip) with additional customized content for targeted fine-mapping. Preliminary analyses have focused on African Americans. We found statistical replication (one-sided p CD58 on chromosome 1, *CD86* on chromosome 3, and an intergenic region on 16q24.1. In all three regions, the most likely causal configuration included two variants (rs12025416 and rs9651076 for *CD58*, rs73179936 and rs73192176 for *CD86*, and rs79034022 and rs8052618 for 16q24.1), with all but one demonstrating posterior inclusion probabilities (PIP) for causality > 0.98 (rs9651076 PIP = 0.83). The two fine-mapped variants for *CD58* were in LD (R2 > 0.2) with the previously identified risk variant (rs1335532) in both Europeans and Africans; however, the fine-mapped variants for the other two loci were not in LD with previously identified variants. This may be because the fine-mapped variants identified for *CD86* were both rare with minor allele frequency

PrgmNr 2825 - Genetic and expression changes associated with recurrence in the perioperative ileal resection period

View session detail

Author Block: **K. Gettler**; Icahn Sch. of Med. at Mount Sinai, NEW YORK, NY
Disclosure Block: **K. Gettler:** None.
Background: Bulk and single cell RNASeq provide complementary mechanistic insight from direct ex-vivo gut sampling in high priority clinical scenarios. Sampling in the perioperative ileal resection period may provide insight into early pathogenic and protective cells, the impact of steroid and anti-TNF use, and allow for integration with genetics in this severe Crohnâ□□s disease cohort. **Methods:** Ribo-depleted bulk RNASeq was performed using paired-end 150 bp reads of 57-90 (mean 70) million reads from 241 terminal ileum pinch biopsy samples. Reads were mapped using STAR, then tested for differential expression between non-recurrence (Rutgeerts scores 0-1, 159 samples) vs. recurrence (Rutgeertâ□□s scores > 2, 82 samples) and those with or without anti-TNF at the time of post-op colonoscopy using DESeq2. DESeq2 normalized expression values of genes corresponding to 84 genes found in an anti-TNF non-response module defined by single cell RNASeq of resected tissues (Martin et al., Cell 2019) were also used to calculate combined Z-scores for samples showing recurrence vs. non-recurrence. Bulk RNASeq differentially expressed genes onto scRNASeq clusters was performed. **Results:** Z-scores were significantly higher among individuals with Rutgeerts recurrence when comparing expression scores of the anti-TNF nonresponse gene module. DESeq2 identified 7,791 genes differentially expressed between recurrence and non-recurrence, with 4,214 genes more highly expressed with recurrence; in contrast, results comparing anti-TNF use were less significant. Upstream regulators of genes upregulated with recurrence included both pro-inflammatory markers such as lipopolysaccharide (p-value = 2.34E-84) and TNF (p-value = 2.89E-69) as well as the anti-inflammatory steroid dexamethasone (p-value = 1.79E-63), while upstream regulators of downregulated genes included HNF4A (p-value = 2.44E-28) and HNF1A (p-value = 1.60E-10). Projecting up-regulated genes onto scRNASeq clusters shows broad gene induction in inflammatory macrophages and activated fibroblasts. **Conclusions:** We correlate a reported anti-TNF non-response gene module with disease recurrence, despite very few differentially expressed genes based on anti-TNF use. We will present developing results from integration of genetic data, including TWAS and eQTL analyses, further refining cellular and molecular mechanisms.

PrgmNr 2826 - Genome-wide analysis of schizophrenia and multiple sclerosis reveals shared genetic loci outside the major histocompatibility complex

View session detail

Author Block: **M. Ahangari**[1], E. Everest[2], T-H. Nguyen[1], E. Turanli[3], B. P. Riley[4]; [1]Virginia Commonwealth Univ., Richmond, VA, [2]Istanbul Technical Univ., Istanbul, Turkey, [3]NIH, Bethesda, MD, [4]Virginia Commonwealth Univ, Richmond, VA

Disclosure Block: **M. Ahangari:** None.

Epidemiological and clinical studies of schizophrenia suggest immune abnormalities in patients with schizophrenia, and genome-wide association studies (GWAS) show enrichment of schizophrenia signals at enhancers that are active in tissues with important immune functions. In this study, we analyzed the largest available GWAS data of schizophrenia (N= 306,011) and multiple sclerosis (MS) (N=41,505), an autoimmune, neurodegenerative disease of the central nervous system, using causal mixture model (MiXeR), and pleiotropy informed false discovery rate (pleioFDR) framework 1) to evaluate the polygenic overlap between schizophrenia and MS, and 2) to improve statistical power for genetic discovery between these disorders, outside the major histocompatibility complex (MHC). We observed polygenic overlap between schizophrenia and MS, and a substantial genetic enrichment in schizophrenia conditional on association with MS and vice versa. By leveraging this cross-trait enrichment, we identified 25 novel schizophrenia risk loci and 21 novel MS loci at conditional FDR

PrgmNr 2827 - Genome-wide association study of juvenile idiopathic arthritis identifies novel risk loci

View session detail

Author Block: **J. Li**[1], Y. Li[1], J. Glessner[1], M. March[1], H-Q. Qu[1], X. Chang[1], M. Khan[1], P. M. Sleiman[2], H. Hakonarson[1], JIA- CHOP consortium; [1]the Children's Hosp. of Philadelphia, Philadelphia, PA, [2]CHOP, Philadelphia, PA

Disclosure Block: **J. Li:** None.

Juvenile Idiopathic Arthritis (JIA) is the most common type of chronic immune-mediated joint disease among children, which can lead to severe complications. JIA is considered to be a highly heterogeneous group of autoimmune conditions, including seven subtypes based on based on clinical presentations. However, overlap in phenotypic presentations and genetic risk loci, as well as altered gene expression common to JIA subtypes, suggest potential shared molecular mechanism underlying JIA subtypes. Therefore, to identify novel genetic risk loci of JIA, we combined the JIA cases of seven subtypes in our study cohort and did a genome-wide association study (GWAS) including 1245 JIA cases and 9250 age, race and sex matched controls. We further performed meta-analysis with a similar combined JIA cohort in GWAS catalog (LÃ³pez-Isac E., Ann Rheum Dis. 2020), which is composed of 3305 patients and 9196 healthy controls. In the meta-analysis, we identified four genome-wide significant novel loci. Functional annotation indicates that these loci overlap with chromatin modification marks, enhancer, promoter regions in immune cell types. The candidate target genes at these loci function in immune regulation. Experimental studies are underway to examine the transcriptional regulation of candidate genes by the noncoding GWAS variants at these loci.

PrgmNr 2828 - Genomic surveillance of SARS-CoV-2 variants for the center of disease control using full viral genome sequencing in 12,000 confirmed patient saliva specimens

View session detail

Author Block: **S. Nahas**[1], C. Bixby[1], C. Goswami[1], M. Keddache[2], M. Sheldon[1], Y. Wang[1], J. Schultz[1], J. McDevitt[1], R. Hager[1], R. Grimwood[1]; [1]Infinity Biologix, Piscataway, NJ, [2]Illumina, San Diego, CA
Disclosure Block: **S. Nahas:** None.
Introduction: To date in the U.S., >30 million COVID-19 cases (~600K deaths) have been confirmed. Genetic variations are known to increase transmission and pathogenicity. The CDC contracted with diagnostic labs to sequence samples across the U.S. The purpose is to perform routine analysis of genetic sequence data to enable the CDC and its public health partners to identify and characterize variant viruses, either new ones identified or those already identified abroad, to investigate how variants impact disease severity and the effectiveness of vaccines and treatment. Infinity Biologix LLC (IBX), was awarded one of these contracts. We sought to highlight the results, from viral genomic sequencing of the first ~12,000 previously confirmed positive COVID-19 patients collected from saliva and interrogate correlations of cycle thresholds (Ct) with genomic sequencing success rates. Methods: Positive RNA patient samples, confirmed with the IBX TaqPath SARS-CoV-2 Assay (EUA200090), were selected for sequencing usingâ¯an N Gene CT value cutoff ofâ¯

PrgmNr 2829 - Identification of miR-eQTLs from small RNA Seq in Children with Asthma

View session detail

Author Block: **A. Tiwari**[1], J. Hecker[1], J. Li[2], A. Kho[3,4], J. CeledÃ³n[5], S. T. Weiss[6], K. G. Tantisira[6,7], M. J. McGeachie[6]; [1]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, Massachusetts, BROOKLINE, MA, [2]Res. Ctr., The Seventh Affiliated Hosp. of Sun Yat-Sen Univ., Shenzhen 518107, China, Shenzen, China, [3]Brigham and Women's Hosp. and Harvard Med. Sch. ,Boston, Massachusetts, BROOKLINE, MA, [4]Computational Hlth.Informatics Program, Boston Children's Hosp., Brookline, MA, [5]Div. of Pediatric Pulmonary Med., UPMC Childrenâ□□s Hosp. of Pittsburgh, Univ. of Pittsburgh, Pittsburgh, PA, USA, Pittsburgh, PA, [6]Brigham & Women's Hosp, Boston, MA, [7]Div. of Pediatric Respiratory Med., Univ. of California San Diego, San Diego, CA, USA, San Diego, CA
Disclosure Block: **A. Tiwari:** None.
Introduction: MiRNAs are increasingly identified as mediators in biological pathways related to asthma and other inflammatory diseases. Identifying genotypes associated with miRNA expression (miR-eQTLs) can assist in understanding the role of microRNA as a mediator in the development of complex traits by providing insights into transcriptional regulation of the microRNAs themselves. **Objective:** The purpose of this study was to identify microRNA expression quantitative trait loci (miR-eQTL) and their association with asthma. **Methods:** We performed small RNA sequencing on 1159 serum samples from the Genetics of Asthma in Costa Rica Study (GACRS), a cross-sectional study of children from of Costa Rica with physician-diagnosed asthma. Small RNA-seq libraries were prepared using the Norgen Biotek Small RNA Lib Prep Kit and sequenced with Illumina NextSeq 500. The exceRpt pipeline was employed for QC, alignment, and annotation of the small RNA-seq data. Whole-genome sequence (WGS) data was available through the TOPMed program. Variant calls were obtained from TOPMed data freeze 8 variant call format files aligned to the GRCh38 genome reference. In our analyses, we included only biallelic SNPs with a minimal depth of coverage of 10 reads that were marked as PASS in the VCF FILTER column. The Matrix eQTL package was used to conduct both cis- and trans- association analysis for each genetic variant passing quality control with each robustly expressed miRNA. Cis- and trans- associations with an FDR **Results:** 1054 GACRS samples passed QC for both miRNA sequencing and WGS. Serum-derived miRNA expression levels of 318 mature miRNAs and 5,734,792 genetic variants passed quality control metrics. We found 25 unique genome variants acting as cis-eQTLs to 36 miRs with p-values ranging from 1.81E-51 to7.19E-06 and 3 miRs with significant trans-eQTLs (p-values 2.79E-13 to 3.15E-11). Among the significant cis-eQTL-miR pairs was rs10175383 located within the primary sequence of hsa-miR-3679, providing a potential mechanism of effect. One miR, hsa-miR-4433b-5p, was previously associated with asthma exacerbations (p-value 0.002), while five cis-eQTL variants were associated with asthma-related phenotypes in the GACRS (p **Conclusion:** We identified a substantial number of miRNAs that are controlled by cis genetic regulatory elements. Our genome-wide miR-eQTL mapping study provides new insights into the genetic regulation of miRNA transcription and the roles of miRNAs using sequencing data, showing that cis-eQTLs may be biologically active variants in asthma.

PrgmNr 2830 - The Genetics of The Antibody Response to Human Adenovirus infection in Healthy Individuals

View session detail

Author Block: **C. Valencia**[1], T. Venkataraman[2], T. Liechti[3], A. V. Valencia-Duarte[4,5], S. Clipman[6], M. Roederer[3], P. Duggal[7,8], B. Larman[9]; [1]Johns Hopkins Sch. of Publ. Hlth., Baltimore, MD, [2]Johns Hopkins Univ.,Immunopathology, Dept. of Pathology, Inst. for Cell Engineering, Baltimore, MD, [3]ImmunoTechnology Section, Vaccine Res. Ctr., NIAID, NHI, Bethesda, MD, [4]Johns Hopkins Univ., Baltimore, MD, [5]Univ. Pontificia Bolivariana, Sch. of Med., Medellin, Colombia, [6]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [7]Johns Hopkins Sch. of Publ. Hlth.Dept. of Epidemiology, Baltimore, MD, [8]Johns Hopkins Sch. of Publ. Hlth.Dept. of Intl. Hlth., Baltimore, MD, [9]Immunopathology, Dept. of Pathology, Inst. for Cell Engineering, Johns Hopkins Univ., Baltimore, MD
Disclosure Block: **C. Valencia:** None.
Mastadenovirus is a viral genus that infect mammals and includes the human adenovirus (HAdv) species that encompasses 7 subgroups with more than 67 serotypes. HAdvs can cause epidemic outbreaks in children, elderly and immunocompromised individuals, however, currently there is no vaccine for these group of individuals. HAdvs can infect different cell types and cause lytic, latent and oncogenic infections. Genetically modified HAdv are also use as vectors for gene therapy, however, pre-existing immunity in patients can limit their use. There is variability in the Immunoglobulin G (IgG) reactivity to the HAdvs capsid proteins, however the genetic contribution to the serum IgG response of HAdv infection remains unknown. Previous studies identified genetic variants in the Human Leukocyte Antigen (HLA) locus associated with IgG levels against Epstein Barr virus, Rubella, and Human polyomavirus. We evaluated the genetics of the antibody response against 52 prevalent HAdvs peptides in a group of 478 European and 147 African ancestry healthy individuals included in the NIH Vaccine Research Center (VRC) study. The mean age was 34 years, and 56% were female. Phage Immunoprecipitaition Sequencing â□□PhIP-seqâ□□ was performed, scaled and dichotomized to classify the samples as positive or negative for reactivity to each peptide. We interrogated 7,331,696 imputed variants across the genome with each of the 52 HAdvs peptides using the penalized quasi-likelihood (PQL) approximation to the GLMM implemented in the R package Genesis. Integrating epitope level antibody reactivity with genetic variation we identified genetic association in the HLA Class II genes with 4 peptides. These peptides are from the minor capsid protein VIII and the core-associated proteins V and VII. The two VIII peptides (amino acids 1-56) represented the same B cell epitope and are from HAdv 48 and HAdv B serotype 3. The other two peptides are from HAdv 55 and correspond to protein V (amino acids 168 - 224) and protein VII (amino acids 141 - 192). HLA fine-mapping identified that the alleles *HLA-DQB1*06:02* (OR=3.79, p=1.85e-8) and *HLA-DRB1*15:01* (OR=3.28, p=7.54e-8) were associated with the antibody reactivity against the minor capsid protein VIII. Conditional analysis for this epitope found an additional association with *HLA-DRB1*11*. In summary, to our knowledge this is the first study that supports the evidence of genetic association of the HLA class-II locus with the IgG antibody response to HAdv infection.

PrgmNr 2831 - Comprehensive genetic screening for variants associated with spermatogenic failure

View session detail

Author Block: **J. Hardy**[1], N. Pollock[1], T. Gingrich[1], T. Banks[1], A. Zielen[1], J. Kuong[1], P. Sweet[1], A. Ramesh[1], A. Basar[1], Z. Nashman[1], H. Jiang[1], K. Hwang[2], J. Vukina[2], K. Orwig[1], S. A. Yatsenko[2], D. Bellisimo[2], A. Rajkovic[3], M. Kurpisz[4], A. N. Yatsenko[5]; [1]Magee Womens Res. Inst., Pittsburgh, PA, [2]Magee Womens Hosp., Pittsburgh, PA, [3]UCSF, San Francisco, CA, [4]Polish Academy of Sci., Poznan, Poland, [5]Univ. of Pittsburgh Magee-Womens Res Inst & Fndn, Pittsburgh, PA

Disclosure Block: **J. Hardy:** None.

Despite significant progress in male infertility research in mouse, no standard clinical genetic assay exists for comprehensively evaluating idiopathic spermatogenic failure (SPGF) in humans. In an effort to establish a clinical test for the routine identification of pathogenic mutations, we assessed genomic variants found in severe forms of SPGF: azoospermia (absence of sperm in the ejaculate) and severe oligozoospermia (TEX11 or novel hemizygous/homozygous losses (e.g. exons 2-8 of *STK11*). Additionally, 20 nonsynonymous, potentially pathogenic SNVs were identified through both sporadic and familial WES studies (n=60). Importantly, a significant proportion of these gene candidates are known to play a role in meiosis (e.g. *BRDT, MEI1, MEIOB, SYCP2,* and *TEX15*). Notably, we were able to detect mosaic chromosomal aneuploidies (e.g. XXY) missed by traditional karyotyping. Our data clearly demonstrates the utility of comprehensive genomic analysis in clinical genetic testing, elucidating molecular etiologies of SPGF at a significantly higher detection rate than existing clinical practice (i.e. aneuploidy and Y chromosome microdeletions). We are currently in process of evaluating the efficiency of an experimentally derived test of more than 150 genes responsible for non-obstructive male infertility with the aim of establishing a single comprehensive clinical genomic screening assay.

PrgmNr 2833 - Glucocorticoid Regulated KCNA5 Mediates Cell Cycle in Fetal Membranes

View session detail

Author Block: **S. J. Cunningham**, T. Allen, L. Feng, T. E. Reddy; Duke Univ., Durham, NC
Disclosure Block: **S.J. Cunningham:** None.
Preterm Premature Rupture of Membranes (PPROM) is a pregnancy complication in which the chorion and amnion weaken and rupture in prior to 37 weeks of pregnancy and before contractions have begun. PPROM results in 30-40% of preterm birth cases and is closely linked to maternal psychological stress. To study the response to glucocorticoid signaling in primary amnion cells, cells from 10 patients were treated with a glucocorticoid receptor agonist and a subset of cells we challenged with Tumor Necrosis Factor Î± in addition. This was repeated with glucocorticoid receptor knocked down and a control siRNA treatment. From this analysis, KCNA5 emerged as a potential GR up-regulated gene with evidence from the literature of having an effect on cell turnover. KCNA5 was directly knocked down in primary amnion epithelial cells from 12 additional patients. KCNA5 knock down and control cells were compared in assays of proliferation and apoptosis. KCNA5 knockdown significantly increased proliferation but seemingly has no effect on apoptosis. To understand to role of increases in KCNA5 expression, KCNA5 was over-expressed in HEK293T cells using a lentiviral system. Over the course of pregnancy, elevated glucocorticoid signaling could increase expression of KCNA5, decreasing proliferation in the amnion epithelium which could lead to membrane weakening and ultimately PPROM.

PrgmNr 2834 - Reprograming human fibroblasts into Sertoli cells: a tool for personalized medicine

View session detail

Author Block: **A. Parivesh**[1], E. Delot[2], A. Reyes[3], J. Ryan[3], S. Bhattacharya[4], V. R. Harley[5], E. J. Vilain[6]; [1]Children s Natl. Med. Ctr., Washington, DC, [2]Children's Natl. Hosp., Washington, DC, [3]Hudson Inst. of Med. Res., Melbourne, Australia, [4]Children s Natl. Hosp., Washington, DC, [5]Hudson Inst. of Med. Res., Clayton, Australia, [6]Children's Natl. Med. Ctr., Washington, DC
Disclosure Block: **A. Parivesh:** None.
Disorders/Differences of Sex Development (DSD) are congenital conditions in which development of chromosomal, gonadal, or anatomical sex is atypical. These achieve poor clinical or research diagnostic outcomes using biochemical assays or *in vitro* cellular/animal model validation of pathogenic variants discovered through DNA sequencing/mapping. The current DSD diagnostic regimen can be augmented by investigating transcriptome/proteome *in vivo*, but it is marred by unavailability of affected gonadal tissue at the disease relevant developmental stage. We try to mitigate this limitation by reprogramming readily available skin tissue derived dermal fibroblasts into Sertoli cells (SC), which could then be deployed for different diagnostic strategies. SCs form the target cell type of choice because they act like organizing center of embryonic gonadal development and many DSD arise when these developmental processes go awry.We employed a computational predictive algorithm for cell conversions called Mogrify to predict the transcription factors (TFs) required for direct reprogramming of human dermal fibroblasts into SCs. We established trans-differentiation culture conditions where stable transgenic expression of these TFs was achieved in 46, XY adult dermal fibroblasts using lentiviral vectors. The resulting Sertoli like cells (SLCs) were validated for SC phenotype using several approaches. These cells exhibited Sertoli like morphological and biophysical properties distinct from fibroblasts as revealed by morphometry and xCelligence assays. They also showed Sertoli-specific expression of molecular markers such as SOX9, PTGDS, BMP4, or DMRT1 as revealed by IF imaging, RNAseq and qPCR. The SLC transcriptome shared ~69% of its differentially expressed genes (DEGs) with a human adult SC transcriptome when compared with fibroblast transcriptome. These cells additionally lacked expression of markers of other gonadal cell types such as Leydig, germ, peritubular myoid or granulosa cells.The trans-differentiation method was applied to a variety of commercially available 46, XY fibroblasts derived from patients with DSD and to normal 46,XX. The resulting cells displayed varying levels of trans-differentiation in comparison to normal 46, XY derived SLCs as estimated by morphological and molecular phenotypes as well as affected molecular pathways, thus showcasing the robustness of this new trans-differentiation model

PrgmNr 2835 - The identification of disease complication risk via disease-disease network: an application to obstetric disorders in the UK Biobank

View session detail

Author Block: **V. Sriram**[1], Y. Nam[2], S. Lee[3], A. Verma[2], D. Kim[2]; [1]The Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Seoul Natl. Univ. Coll. of Med., Seoul, Korea, Republic of

Disclosure Block: **V. Sriram:** None.

Recent studies have shown that women with pregnancy-related disorders are at risk for a variety of long-term complications. However, the underlying pathophysiology of these associations remains undetermined, and it is unclear how much of a role is played by genetics. Thus, the interactions between obstetric diseases and subsequent phenotypes suggest that a network-based view that incorporates knowledge of other diseases and genetic associations will enable us to uncover the exact basis of pregnancy-related disease complications. A disease-disease network (DDN), where nodes represent diseases and edges represent commonalities between diseases such as shared SNPs, can be used to evaluate genetic correlations between phenotypes. By applying graph-based semi-supervised learning (GSSL), an approach for topology-based label propagation, potential shared genetic associations between obstetric disorders and their disease complications can be identified. We constructed a SNP-based DDN from UK Biobank (UKBB) PheWAS summary data, which included 232 phenotypes for 27,193 imputed variants. To select the most appropriate obstetric phenotypes for GSSL analysis, we applied case count and weighted degree thresholds. A minimum case count of 500 was used to ensure sufficient statistical power for the PheWAS summary results. Then, phenotypes were sorted according to normalized weighted degree in order to select those that shared the highest number of SNPs with other phenotypes. The final obstetric disorders we considered included hemorrhage during pregnancy, early labor, placenta previa, antepartum hemorrhage, and hemorrhage in early pregnancy. For each obstetric phenotype under consideration, we assigned a â□□+1â□□ label to the disease and performed GSSL label propagation. The results of our complication risk prediction for placenta previa and subsequent phenotypes yield AUCs ranging from 0.604 to 0.678 when compared to clinical co-occurrences derived from UKBB electronic health records, suggesting that genetic associations play varying roles in disease complications. This current result suggests that our methodology holds promise as a clinical tool for the identification of possible genetic drivers for the development of disease complications.

PrgmNr 2836 - Genome-wide association study and gene-based analysis of participants with hemophilia A and inhibitors in the My Life Our Future study

View session detail

Author Block: **S. Lessard**[1], C. He[2], D. Rajpal[1], K. Klinger[1], C. Loh[3,4], T. Harris[4,5], J. Dumont[6];
[1]Translational Sci., Sanofi, Framingham, MA, [2]American Thrombosis and Hemostasis Network, Rochester, NY, [3]Nimbus Therapeutics, Cambridge, MA, [4]Employee of Bioverativ, a Sanofi company, at the time of study, Cambridge, MA, [5]Repertoire Immune Med.s, Cambridge, MA, [6]Sanofi, Cambridge, MA
Disclosure Block: **S. Lessard:** Major Stockholder/Ownership Interest; Sanofi. Salary/Employment; Sanofi.
Up to 30% of people with hemophilia A (PwHA) develop factor VIII (FVIII) inhibitors, the most serious complication of hemophilia A treatment, as factor replacement is made ineffective. The mechanism of inhibitor development is poorly understood. The My Life Our Future (MLOF) repository gathered *F8* and *F9* gene mutation information, phenotypic data, and biological material from >11,000 PwHA and hemophilia B enrolled at US hemophilia treatment centers from 2012 to 2018, including >5000 whole-genome sequences (WGS). Identifying inhibitor-associated genes may contribute to understanding why patients develop inhibitors.
We performed a genome-wide association study of inhibitor and gene-based analyses to identify inhibitor-associated genes in male PwHA from the MLOF repository. Those of reported active inhibitors and prescribed bypassing agents or immune tolerance induction or with a history of inhibitors were included (>5 Bethesda units at peak were considered high-titer inhibitors). Non-inhibitor PwHA were defined as controls (no history of inhibitors or evidence of active inhibitors). Those of unknown inhibitor status were excluded. We performed discovery association analyses (single variant and gene-based) in participants of European ancestry and attempted to replicate significant findings in those of African and Hispanic ancestry.
We identified a genome-wide significant association with 6:32438468_CA/C (P=4Ã□10^{-8}), located near *HLA-DRA* in the human leukocyte antigen (HLA) locus, in PwHA with *F8* intronic inversions. Low resolution WGS-based HLA typing showed independent associations with HLA alleles HLA DRB1*15:01 and DQB1*03:03 post Bonferroni correction. Variant aggregation tests identified low-frequency coding variants (minor allele frequency GRID2IP, encoding the GRID2-interacting protein, that were significantly associated with high-titer inhibitors in participants of European ancestry (P=7Ã□10^{-7}). We replicated this finding in participants of African ancestry (P=0.008).
Our study confirms association of DRB1*15:01 with FVIII inhibitors and reveals a novel association with HLA-DQB1*03:03 in PwHA with intronic inversions of *F8*. The results also implicate *GRID2IP* with the presence of inhibitors, suggesting a new role of this gene in autoimmunity.

PrgmNr 2837 - Integrative analyses of germline variation and somatic mutations identify germline-somatic interactions in myelofibrosis susceptibility

View session detail

Author Block: **D. Brown**[1], Y. Wang[1], A. St. Martin[2], S. R. Spellman[3], S-H. Lin[1], O. Lee[1], A. Klein[1], T. Zhang[1], W. Zhou[4], K. Jones[4], A. Hutchinson[4], J. Liu[4], B. Zhu[4], C. Dagnall[4], S. Khan[1], M. Rafati[1], T. Wang[5], S. J. Lee[6], M. Yeager[4], S. A. Savage[1], W. Saber[2], S. M. Gadalla[1], M. J. Machiela[1]; [1]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD, [2]Ctr. for Intl. Blood and Marrow Transplant Res., Med. Coll. of Wisconsin, Milwaukee, WI, [3]Ctr. for Intl. Blood and Marrow Transplant Res., Natl. Marrow Donor Program, Minneapolis, MN, [4]Cancer Genomics Res. Lab., Frederick Natl. Lab., Frederick, MD, [5]Ctr. for Intl. Blood and Marrow Transplant Res., Div. of Biostatistics, Med. Coll. of Wisconsin, Milwaukee, WI, [6]Ctr. for Intl. Blood and Marrow Transplant Res., Med. Coll. of Wisconsin; Clinical Res. Div., Fred Hutchinson Cancer Res. Ctr., Seattle, WA

Disclosure Block: **D. Brown:** None.

Myelofibrosis (MF) is a rare myeloproliferative neoplasm (MPN) characterized by bone marrow fibrosis, progressive bone marrow failure, and increased risk of acute myeloid leukemia. While myelofibrosis arises from somatic driver mutations in *JAK2*, *MPL*, and *CALR*, previous work on MPNs suggests a heritable component. To comprehensively examine the genetic etiology of MF, we performed the first integrative analysis of SNP array genotyping (using Infinium Global Screening Array), and telomere length (TL, using qPCR assay). Somatic mosaic chromosomal alterations (mCAs) were called with MoChA using raw genotyping intensity data to detect allelic imbalances. The study included 937 patients who received hematopoietic cell transplant for MF between 2000 and 2016 and who had an available pre-HCT blood sample at the Center for International Blood and Marrow Transplant Research Repository. GWAS of 827 European MF cases and 4,135 genetically-matched controls identified six independent loci at genome-wide significance (P-8); four of which replicate prior MPN susceptibility loci [9p24.1(*JAK2*), 5p15.33(*TERT*), 3q25.33(*IFT80*), and 4q24(*TET2*)] and two novel MF loci [6p21.35(*HLA-DQB1-AS1*) and 17p13.1(*TP53*)]. A transcriptome-wide association study using whole blood GTEx data highlighted the 9p24.1 locus with increased *JAK2* expression associated with elevated risk of MF (P= 2.18Ã□10^{-19}). A strong statistical colocalization further indicated shared genetic etiology between eQTL and this *JAK2* locus (HyPrColoc Posterior Probability= 0.6). Utilizing available clinical mutation data on a subset of 185 patients, MF cases carrying the risk haplotype of the 9p24.1(*JAK2*) susceptibility locus were observed to more frequently have the $JAK2^{V617F}$ mutation (P= 0.02). Detectable autosomal mCAs were abundant in MF cases with 67.4% having at least one mCA (compared to ~3% in the general population) and 27.6% having an mCA spanning *JAK2* (mostly copy number neutral loss of heterozygosity (CNLOH)). In addition, a *cis* relationship was identified at 9p24.1 in which the MF risk haplotype was predominantly duplicated by CNLOH (binomial P=1.36Ã□10^{-9}). Finally, we observed an inverse association between autosomal mCAs and telomere length (OR= 0.22, 95% CI= 0.07-0.65, P= 6.40Ã□10^{-3}). Our results suggest a molecular framework for the genetic etiology of MF in which germline variation at *JAK2* predisposes to the acquisition of a somatic $JAK2^{V617F}$ mutation and subsequent duplication of $JAK2^{V617F}$ by mCAs (usually CNLOH). This process leads to aberrant *JAK2* activity and increased clonal proliferation, exhausting telomere length and increasing genomic instability in MF patients.

PrgmNr 2838 - *MAPT* Structural Haplotypes in Chronic Traumatic. Encephalopathy

View session detail

Author Block: **X. Han**[1,2], J. Chung[3], K. Atherton[1], M. M. Khan[1], C. Shea[1], E. Nair[4], Z. H. Baucom[5], B. Abdolmohammadi[4], M. Uretsky[4], B. M. Martin[6,7], J. N. Palmisano[6,7], K. Farrell[8,9,10], J. J. Farrell,[3], J. D. Cherry[4,7,11,12,13], V. E. Alvarez[4,7,11,12], B. R. Huber[4,7,11,12,2], M. L. Alosco[4,7,2], K. L. Lunetta[5], Y. Tripodis[4,5,7], T. D. Stein[4,7,11,12,13], L. A. Farrer[3,5,7,2], J. F. Crary[8,9,10], A. C. McKee[4,7,11,12,13], J. Mez[4,7,2]; [1]Bioinformatics Graduate Program, Boston Univ., Boston, MA, [2]Dept. of Neurology, Boston Univ. Sch. of Med., Boston, MA, [3]Section of BioMed. Genetics, Dept. of Med., Boston Univ. Sch. of Med., Boston, MA, [4]Boston Univ. Chronic Traumatic Encephalopathy Ctr., Boston Univ. Sch. of Med., Boston, MA, [5]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [6]Biostatistics and Epidemiology Data Analytics Ctr., Boston Univ. Sch. of Publ. Hlth., Boston, MA, [7]Boston Univ. Alzheimer's Disease Res. Ctr., Boston Univ. Sch. of Med., Boston, MA, [8]Dept. of Pathology, Icahn Sch. of Med. at Mount Sinai, New York, NY, [9]Fishberg Dept. of NeuroSci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [10]The Ronald M. Loeb Ctr. for Alzheimerâ□□s Disease, Icahn Sch. of Med. at Mount Sinai, New York, NY, [11]VA Boston Hlth.care System, Boston, MA, [12]Dept. of VA Med. Ctr., Bedford, MA, [13]Dept. of Pathology and Lab. Med., Boston Univ. Sch. of Med., Boston, MA

Disclosure Block: **X. Han:** None.

Introduction: Chronic traumatic encephalopathy (CTE) is a neurodegenerative tauopathy associated with repetitive head impact (RHI) exposure. Variation in the *MAPT* gene, which encodes the tau protein, is associated with other tauopathies, but its role in CTE remains unclear. The 17q21 region, including *MAPT,* contains a megabase-long inversion polymorphism (H1/H2, unique to European ancestry) and many copy number variations, including Î±, Î² and Î³ segments, which determine various structural subhaplotypes. *Boettger et al.* previously showed that these structural subhaplotypes can be imputed with SNPs from a genome-wide array and a reference haplotype panel. Here, we imputed 17q21 structural haplotypes and tested their associations with CTE neuropathological endophenotypes. **Methods:** 144 men with European ancestry and RHI exposure (117 neuropathologically-confirmed CTE cases and 27 controls) were genotyped on 4,739 SNPs across the *MAPT* region. Using the *Boettger et al.* reference haplotype panel, SHAPEIT and IMPUTE2, 12 biallelic surrogate markers for H1/H2, Î±/Î²/Î³ segments and 5 common subhaplotypes (H1Î²1Î³1, H1Î²1Î³2, H1Î²1Î³3, and H2Î±2Î³2) were imputed. We tested associations between these haplotypes and CTE endophenotypes [CTE status, CTE stage (0-4) and semi-quantitative measures of tau burden (0-3) in eleven brain regions affected in CTE] in logistic and linear regression models adjusted for age and principal components of population substructure. **Results:** The 12 surrogate markers were well-imputed (imputation quality R^2: 0.57-0.74). The H1 haplotype (freq = 0.85) was not associated with CTE status or stage, but was significantly associated with tau burden in the amygdala in the full sample ($\hat{I}^2$ = 0.32; P = 0.049) and among CTE cases ($\hat{I}^2$ = 0.58; P = 0.002). The H1Î²1Î³1 subhaplotype (freq = 0.37) was significantly associated with tau burden in the amygdala ($\hat{I}^2$ = 0.25; P = 0.049) among the CTE cases. The H1Î²1Î³1 subhaplotype was also significantly associated with CTE stage in the full sample ($\hat{I}^2$ = 0.33; P = 0.03) and among CTE cases ($\hat{I}^2$ = 0.27; P = 0.03). Additionally, among H1 homozygotes, the numbers of Î³ segments, which is physically located at the upstream of the gene KANSL1, had a significant protective effect on CTE stage ($\hat{I}^2$ = -0.41; P = 0.01). **Conclusion:** Imputation of structural haplotypes with the SNP/haplotype reference panel allowed for new insights into the relationship between structural variants in the *MAPT* region and CTE, including a significant association between the H1 haplotype and amygdala tau burden and the H1Î²1Î³1 subhaplotype, CTE stage and amygdala tau burden.

PrgmNr 2839 - Autism spectrum disorder phenotype subset in attention deficit hyperactivity disorder: An analysis of first-tier test results and phenotype spectrum and severity

View session detail

Author Block: **J. J. Kapalanga**[1,2], R. Kamran[2], T. M. Parikh[1]; [1]Western Univ.GBHS, London, ON, Canada, [2]McMaster Univ., Hamilton, ON, Canada
Disclosure Block: **J.J. Kapalanga:** None.
The autism spectrum disorder (ASD) phenotype is frequently identified in a subset of individuals diagnosed with attention deficit hyperactivity disorder (ADHD). Both ASD and ADHD are highly heritable and share a phenotype distributed across five domains including adaptive malfunctioning, social competence deficits, intellectual deficit/developmental delay, emotional dysregulation, and a history of medical disorders. Individuals with ASD or ADHD can present with a combination of hyperactivity, impulsivity, anxiety, mood issues, developmental delay/cognitive deficits, social dysfunction, gastroesophageal reflux, feeding difficulties during infancy and hypermobility syndrome. **Objectives:** To analyze first tier genetic test results and phenotype spectrum and severity in a cohort of children with an initial diagnosis of ADHD or suspected ASD, aged 6 to 17 years who have been followed for at least 3 years. To compare first tier test results, phenotype spectrum and severity between ADHD and a subset of individuals diagnosed with ASD. **Methods:** 967 children with suspected ADHD and or ASD were referred from the community through primary care providers between January 2010 to December 2020. Children with at least one phenotypic feature from each of five phenotype domains were identified. The eligible children were assessed with instruments used for assessing ASD and/or ADHD. Based on these assessments children were diagnosed as ASD and ADHD. Symptoms were quantified as 1 (mild), 2 (moderate) and 3(severe). Following diagnosis, first tier genetic testing were performed. Multinomial logistic regression models were used to explore the association of phenotype spectrum and phenotype severity, and ASD and ADHD. Multivariate analysis was used to compare phenotype severity and first-tier test results between ASD and ADHD. **Results:** Of the 967 children 397 children fulfilled inclusion criteria. Among the 397, 250 (63%) were male and 147 (37%) were female. Median age at diagnosis for males was 5.8 years and for females, 7 years. Among the 397 children 64 (16%) were diagnosed with ASD and 333 (84%) with ADHD. The overall phenotype severity was .78 (95% CI, 0.81-0.89) for ADHD and .92 (95% CI, 0.82-0.88) for ASD. In the ADHD group 47 (14 %) and in the ASD group 10 (16.2 %) had abnormal first tier test results. **Conclusion**: The findings suggest that ASD and ADHD share a phenotype spectrum, but phenotype severity is significantly greater in ASD. The results of first tier test results were similar for ASD and ADHD with no obvious disorder specific abnormal test result, consistent with findings in previous studies.

PrgmNr 2840 - Brain trans-pQTL analysis of GWAS loci reveals novel link between psychiatric disorders and neurodegenerative diseases

View session detail

Author Block: **T. S. Wingo**[1], S. M. Vattathil[1], E. S. Gerasimov[1], Y. Liu[1], D. A. Bennett[2], N. T. Seyfried[1], A. I. Levey[1], A. P. Wingo[1]; [1]Emory Univ. Sch. of Med., Atlanta, GA, [2]Rush Univ., Chicago, IL
Disclosure Block: **T.S. Wingo:** None.
Background: Many psychiatric disorders and neurodegenerative diseases have shared symptomatology and genetic basis. GWAS of each of these brain illnesses identified multiple genetic loci associated with each illness; however, how these loci confer disease risk remains unclear. Here, we hypothesize that some genetic risk is explained by influences on brain protein expression. To test this hypothesis, we examined whether GWAS loci were associated with either proximally or distally encoded proteins in the human brain. Identified proteins were then tested for evidence of causality, and finally, results were combined to identify shared proteins linking psychiatric disorders and neurodegenerative diseases.
Methods: The following brain illnesses were considered: Alzheimer's disease (AD), frontotemporal dementia, amyotrophic lateral sclerosis, Parkinson's disease (PD), insomnia, depression, schizophrenia, anxiety disorders, bipolar disorder, post-traumatic stress disorder, problematic alcohol use, and neuroticism. Independent GWAS signals from these conditions were identified and tested for association with brain proteins in a discovery dataset (N=503) of human brain proteomes. Significant results from the discovery analysis were followed-up in an independent replication dataset (N=135) of human brain proteomes. Next, proteins with consistent associations were further examined using summary data-based Mendelian randomization and HEIDI to test for evidence of causality/pleiotropy and exclude associations likely due to linkage disequilibrium, respectively. Results were combined across all conditions to examine for evidence of shared proteins or convergence of physical protein-protein interaction using direct interaction data from BioGRID database.
Results: Among the GWAS significant loci, we identified 19 pairs of trans-pQTLs and the distal protein it regulates in AD, PD, schizophrenia, alcohol problematic use, neuroticism, and depression. Fifteen pairs showed evidence that the protein mediates the association between the GWAS SNP and disease after Mendelian randomization and HEIDI. We found that these 15 proteins had evidence of substantial protein-protein interactions. We also found evidence linking AD, neuroticism, and depression by syntaxins, syntaxin-binding proteins and secretory membrane proteins.
Conclusion: These networks of interacting proteins reveal novel links and biological insights between psychiatric and neurodegenerative diseases.

PrgmNr 2841 - Copy number variation identification and association study on 3,800 Alzheimer's disease whole genome sequencing data from the Alzheimer's Disease Sequencing Project (ADSP)

View session detail

Author Block: **A. Tucci**[1], M. Conery[2], O. Valladares[2,3], Y-F. Chou[2,3], Y. Leung[3,4], A. Kuzma[3,4], W. Lu[5], G. D. Schellenberg[2], L-S. Wang[3,4], J-Y. Tzeng[5], W-P. Lee[2]; [1]Bioinformatics Res. Ctr., North Carolina State Univ., Raleigh, NC, [2]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [3]Inst. for BioMed. Informatics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [4]Penn Neurodegeneration Genomics Ctr., Philadelphia, PA, [5]North Carolina State Univ. Dept. of Statistics, Raleigh, NC
Disclosure Block: **A. Tucci:** None.
Background: Estimation of heritability of Alzheimerâ□□s Disease (AD) ranges from 49-79%, but the conventional single nucleotide variants (SNVs) identified to date account for Method: To accelerate the discovery of genetic traits of AD, the Alzheimerâ□□s Disease Sequencing Project (ADSP) sequenced whole genomes of AD cases and cognitively normal controls from multiple ethnic groups. Leveraging this large-scale Whole Genome Sequence (WGS) collection, we conducted AD genetic association analyses across a full spectrum of CNVs (small or large, common or rare, and coding or non-coding in genomic regions). We developed a bioinformatics pipeline consisting of four steps. First, we applied three CNV calling algorithms, CNVnator, JAX-CNV, and Smoove, on each sample. Results from each tool were then merged by Svimmer. Second, we employed GraphTyper for genotyping. GraphTyper is a graph-pangenome-based method that may mitigate bias of human references. Third, we resolved issues of multiple CNVs overlapping in a region with bedtools, which is a requirement for working with most association study tools. Finally, with the CNV callsets, we performed genome-wide association analyses on common and rare CNVs to identify CNV regions that contribute to AD. For common CNVs, we defined CNV regions (CNVRs) using a density-threshold trimming procedure (as defined in CNVRuler) and conducted CNVR-based analysis. For rare CNVs, we preformed chromosome-wide CNV collapsing tests. In all association analyses, we adjusted for age, sex, APOE e4 dose and race-ethnicity status.
Result: The ADSP dataset consists of 1,737 AD cases and 2,063 cognitively normal controls. For each sample, we identified an average of 7,959 CNVs by three algorithms. After merging CNVs of all samples, we obtained 56,316 deletions and 16,390 duplications for GraphTyper joint genotyping. For common association analysis, there were ~30 CNVRs significantly associated with AD at genome-wide significance level. Rare CNV collapsing tests revealed a significant signal on 17p13, and rare CNV burden tests showed signals on chromosomes 1 and 5. Further validations of those CNVs are under process.
Conclusion: We developed a scalable CNV detection pipeline and applied it to 3,800 ADSP WGS samples. The preliminary results of the association study validate some known AD, dementia, and dystonia genes. To the best of our knowledge, this is the first large-scale CNV investigation of AD using WGS data.

PrgmNr 2842 - Exome Sequencing Identifies Variants on the X-Chromosome Associated with Alzheimerâ□□s Disease in the Alzheimerâ□□s Disease Sequencing Project

View session detail

Author Block: **J. Chung**[1], X. Sun[2], M. Schmidt[3], Y. E. Song[4,5], M. Grunin[4], E. Palmer[4], W. S. Bush[4], A. C. Naj[6], F. RAJABLI[3], K. L. Hamilton-Nelson[3], X. Zhang[1,7], Y. Leung[6], L-S. Wang[6], R. Mayeux[8], G. D. Schellenberg[9], L. A. Farrer[1,10,11,7,12], M. A. Pericak-Vance[3], J. L. Haines[4], B. Kunkle[3], E. R. Martin[3]; [1]Dept. of Med., Boston Univ. Sch. of Med., Boston, MA, [2]Dept. of Med., Boston Univ. Sch. of Med., Boston, MA, United States, Boston, MA, [3]John P. Hussman Inst. for Human Genomics, Univ. of Miami, Miami, FL, [4]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [5]Cleveland Inst. for Computational Biology, Case Western Reserve Univ.,, Cleveland, OH, [6]Dept. of Pathology and Lab. Med., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [7]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [8]Dept. of Neurology, Columbia Univ., New York, NY, [9]Univ Pennsylvania Sch Med, Philadelphia, PA, [10]Dept. of Neurology, Boston Univ. Sch. of Med., Boston, MA, [11]Dept. of Ophthalmology, Boston Univ. Sch. of Med., Boston, MA, [12]Dept. of Epidemiology, Boston Univ. Sch. of Publ. Hlth., Boston, MA

Disclosure Block: **J. Chung:** None.

Background: AD is the most frequent neurodegenerative disease with a strong genetic component. GWAS studies have identified numerous associated loci, but have been focused only on the autosomes. We aimed to identify functional rare variants on X-Chromosome (Chr), which contribute to the AD risk using the whole exome sequencing (WES) of a multi-ethnic AD sample. **Methods:** We analyzed X-Chr data from the WES Release 2 dataset in the ADSP, which were jointly-called across 9 multiethnic datasets of 11,460 non-Hispanic White (NHW), 4,108 African American, and 2,159 Hispanic samples. Variants on X-Chr of 8,789 AD cases and 9,613 controls were quality-controlled (QC) within each source of dataset (i.e. subset) and capture kit and were functionally annotated using Variant Effect Predictor (VEP). After QC, 494 nonsynonymous variants with severe pathogenicity prediction and presence in more than two subsets were selected and further analyzed within each sex and ethnic group. A study-wide significance threshold was assigned at $1.01x10^{-4}$ (= 0.05 / 494). We performed AD association tests using logistic regression, accounting for age, relationship matrix, principal components for population substructure, and subsets. Phenomic impact in carriers of selected variants was examined using Geno2MP database. **Results:** In the 6,908 NHW women (3,684 cases / 3,224 controls), we identified one study-wide significant association with AD at rs73227155 (minor allele counts = 84, OR = 2.61, P = $8.3x10^{-5}$), which was present across all subsets. Rs73227155 is a rare, missense variant located in MTMR8 (allele frequency in non-Finish Europeans from gnomAD = 0.006) and predicted to be deleterious by SIFT in VEP. However, rs73227155 was not associated with AD in the 4,552 NHW men (2,580 cases / 1,972 controls; OR = 1.25, P = 0.27) and other ethnic samples (P > 0.05). No other significant variant was found across other ethnicities in sex-specific manner. From the meta-analysis across sexes and ethnic groups, we did not identify other significant associations. From Geno2MP, we found 80 carriers of the minor allele of rs73227155 or their relatives, and 30 subjects out of 80 expressed phenotypic abnormalities in nervous system including intellectual disability, cerebellar hypoplasia, and autism. **Conclusion:** We identified one highly functional rare missense variant, rs73227155, in *MTMR8*, on X-Chr among the NHW women. MTMR8 functions in autophagy in neuronal cells for removing lipids and amyloids. Replication studies will use independent subjects from an AD GWAS sample after TOPMed-imputation in AD Genetics Consortium (ADGC) and Whole Genome Sequencing of ~17,000 multiethnic sample in ADSP.

PrgmNr 2843 - Genetic Variant, rs2304672, in 5â UTR region in *hPER2* gene and its association to Parasomnias and other sleep disorders in Puerto Rico

View session detail

Author Block: **G. Pagan**; Univ. de Puerto Rico, Recinto de Mayaguez, Mayaguez, PR
Disclosure Block: **G. Pagan:** None.
Sleeping is one of the main physiological processes of cellular and physiological restoration. Inadequate sleep patterns can impede adequate systemic restoration and can lead to the development of chronic health conditions. In 2014, both World Health Organization and the Central Disease Control in the United States categorized inadequate sleep patterns as a public health crisis. Both entities majorly agree that, approximately 40 percent of the general population suffers from sleep disorders, with 60 percent of these opt for medication to induce sleep. Researchers from Epidemiologist suggest that a series of external factors, have been theorized to contribute to lack in sleep quality. Including high levels of stress, sounds, and light quality. Alternatively, there has been a limitation regarding studying internal factors affecting sleep quality and sleep behavior. Parasomnia is an abnormal sleep behavior in any of the five sleep stages characterized by sleep walking, sleep talking, bedwetting, nightmare disorders, and sleep paralysis principally. To date, little is known about this sleep abnormality and its origins. In this study, we referenced data from the 1000 Genome Project to study the genetic variant, rs2304672, of the *hPER2* gene in Puerto Rico, whose general population houses the highest percentage of the variant in all the Americaâs. Provided data suggests that this genetic variant is linked to poor sleep quality and parasomnias. In Puerto Rico, modern demographics evidence that the population is comprised of three main ethnic groups. We theorize that a bottleneck effect is an influential factor in the distribution and prevalence of said genetic variant in the island. Our results found that the genetic variant is common in the southwestern and southeastern coastal regions of Puerto Rico. Additionally, we found that the genetic variant frequency in the entire island is of approximately 4 percent in the general population of Puerto Rico.

PrgmNr 2844 - Predicting Mortality Among Ischemic Stroke Patients Using Polygenic Risk Scores Derived From the Disease-Related Biological Pathways

View session detail

Author Block: **J. Li**[1], D. Chaudhary[1], S. Shahjouei[1], P. Hendrix[1], Regeneron Genetics Center, C. Griessenauer[1], D. J. Carey[2], R. Zand[3], V. Abedi[1]; [1]Geisinger Med. Ctr., Danville, PA, [2]Geisinger Clinic, Danville, PA, [3]Geisinger, Danville, PA

Disclosure Block: **J. Li:** None.

Objective: Stroke is the second leading cause of death worldwide. Our previous study (Li, J 2021 Neurology Genetics) identified pathway-specific polygenic risk scores (PRS)s associated with ischemic stroke (IS). The goal of this study is to determine whether these disease-related PRSs are also associated with overall survival post-IS and to develop PRS-based mortality prediction models. Methods: 1756 patients with acute IS and European ancestry were identified by leveraging Electronic Health Records and chart review confirmation. The cohort was split into discovery (n=1226) and replication (n=530) with up to 3-year post-IS observation. PRS derived from previously identified gene sets of Gene Ontology (GO) Biological Process was used to stratify the cohort. Univariate Cox proportional hazards (Coxph) regression analysis was used for primary screening of prognostic PRSs. Only the significantly associated PRSs were used to construct a multivariate Coxph model. PRS selection was conducted by LASSO penalized cox regression model from the training cohort. The selected PRSs and their coefficients were applied to the testing cohort. Results: Using all common or low-frequency variants (0.01

PrgmNr 2845 - Replicating genetic associations in Alzheimer's Disease using data from Eisai's BAN2401 ABBA 201 clinical trial

View session detail

Author Block: **F. Tao**, C. Zhang, A. A. Mitchell, M. Nagle, J. E. Hutz; Eisai Ctr. for Genetics Guided Dementia Discovery, Cambridge, MA
Disclosure Block: **F. Tao:** Salary/Employment; Eisai Inc..
Introduction: Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by memory loss and cognitive impairment. ABBA 201 is a randomized, controlled clinical trial to determine whether BAN2401, an anti-amyloid Î² antibody, can slow the progression of early AD. Over 800 amyloid positive subjects with mild cognitive impairment or early AD (Mini-Mental State Exam MMSE >= 22) participated in the clinical trial, and 665 subjects consented to participate in exploratory genetic research. The aim of the genetic study is to replicate or further characterize known genetic associations in AD phenotypes. **Methods:** DNA samples from 665 participants (335 males, 330 females) were genotyped using Illumina Infinium Multi-Ethnic Global-8 v1.0 BeadChip. The clean dataset after standard quality control includes 47 million genotyped and imputed variants in 641 European-ancestry individuals. Variants in about 90 genes previously reported in AD were selected for analysis. Phenotypes data include age at onset (AAO) of AD, baseline cognitive scores and laboratory tests. Genetic association analysis was performed in PLINK v1.9, adjusting for age, sex, and top five principal components. *APOE* E4 carrier status, a known AD risk variant, was intended to serve as a positive control to confirm its effect through genome-wide analysis, and therefore was not included as a covariate. Other relevant factors such as disease duration were included as covariates in the analysis as appropriate. **Results:** The genetic analyses validated two previously known AD-associated loci: (1) *APOE* was confirmed as the strongest locus in association with AAO of AD (beta = -2.4 years, P = 5.5E-7 for APOE E4 allele, which consists of rs429358-C and rs7412-C). Rs429358 is associated with AAO with beta = -2.5 years and P = 3.2E-7. Rs7412 is not associated with AAO (P=0.29). (2) We found a suggestive association signal in the *FCER1G* gene on chromosome 1 with AD Composite Score (ADCOMS) (beta = -0.06, P = 7.2E-6 for top SNP rs1136207 T allele). The top SNP rs1136207 was also associated with Clinical Dementia Rating - sum of boxes (CDR-SB) (beta = -0.47, P = 1.8E-5). Genetic association in the *ADAMTS4*/*FCER1G* locus was previously reported in the AD GWAS by Jansen *et al.* (PMID 30617256). In addition to *FCER1G*, other candidate genes in this region include *NDUFS2*, *ADAMTS4*, *TOMM40L*, and *APOA*. **Conclusions:** This study confirms that *APOE* E4 is associated with earlier AAO of AD. We also replicated the association of the *ADAMTS4*/*FCER1G* locus previously reported in AD GWAS. Further study will expand the scope to genome-wide association analysis to identify potential novel genetic associations in AD phenotypes.

PrgmNr 2846 - Severity of dementia ratings in Hispanic Latinos as a function of country of origin and APOE

View session detail

Author Block: **P. Mena**[1], A. Zaman[1], M. Prough[1], S. Tejada[1], M. Contreras[1], L. D. Adams[2], C. Silva[3], M. Illanes-Manrique[4], M. R. Cornejo-Olivas[5], B. Feliciano[6], J. M. Vance[1], M. A. Pericak-Vance[7], M. L. Cuccaro[8]; [1]Univ. of Miami, Miami, FL, [2]Univ Miami, Miami, FL, [3]Univ. Central del Caribe, Bayamon, Puerto Rico, [4]Neurogenetics Res. Ctr., Inst. Natl. de Ciencias NeurolÃ³gicas, Lima, Peru, [5]Inst. Natl. de Ciencias Neurologicas, Lima, Lima, Peru, [6]PRADI, UCC, SJB Sch. of Med., Caguas, PR, [7]Miami, FL, [8]Univ Miami Sch Med, Miami, FL

Disclosure Block: **P. Mena:** None.

Background: Hispanic Latinos are underrepresented in Alzheimer disease (AD) research. As a group, Hispanic Latinos are culturally and genetically diverse. A source of diversity is country of origin which confers differences in symptoms as well as sociocultural and genetic backgrounds. While differences in symptoms may arise from genetic background, sociocultural differences may affect reporting of symptoms in different countries. For this study we investigated the relationship of country of origin and *APOE* genotype to dementia severity ratings in a multi-country of origin dataset. We hypothesize that there will be differences in severity related to country of origin. **Methods:** Participants (N=353) whose countries of origin were Puerto Rico (N=242), Cuba (N=29), and Peru (N=83) were drawn from genetic studies of AD in Hispanic Latino populations. Participants underwent standard clinical assessments that included the Clinical Dementia Rating scale (CDR). We restricted participants to those with CDR scores of $\geq$0.5. The CDR box score sum (CDR-sb; 0-18) was used as an outcome measure; sex, *APOE* (*e*4 present/*e*4 absent), age at onset (AAO), age at exam (AAE), and country of origin were predictors. We tested the association of these predictors to CDR-sb using linear regression. **Results:** Our sample was 67% female with a mean AAE of 77.8 (sd=9.4) years and mean AAO was 72.6 (sd=9.6); 42% had an *e*4 allele. The mean CDR-sb was 9.4 (sd=6.0). Our linear regression showed that compared to Puerto Ricans, Cubans had lower CDR-sb scores (B= -3.65, p B=1.82, p = 0.008). We also found that AAE was positively associated with CDR-sb (B=0.51, pB=-0.33, p B=0.04, p=0.95) or APOE (B=0.89, p = 0.12) were significant predictors of CDR-sb scores. **Discussion:** Our results showed that country of origin was a significant predictor of dementia severity ratings with Peruvians having the highest score and Cubans the lowest. Not surprisingly lower age at onset and greater age at exam were significant contributors to ratings of dementia severity. However, the presence of *APOE e*4 did not contribute significantly to dementia severity ratings. Given that ratings of dementia severity are used to classify participants for genetic studies of AD/dementia it is critical to understand the role of sociocultural influences on these ratings in studies that enroll participants from different countries. Future studies should investigate population specific sociocultural and genetic contributions to dementia severity ratings and harmonize measures such as the CDR across populations.

PrgmNr 2847 - SnRNA-seq probing of the differential vulnerability of motor neurons in ALS

View session detail

Author Block: **P. Alipour**[1,2], J. Ross[1,2], D. Rochefort[1,2], P. Dion[1,2], G. A. Rouleau[1,2,3]; [1]Montreal Neurological Inst., and Hosp., McGill Univ., Montreal, QC, Canada, [2]Dept. of Human Genetics, McGill Univ., Montreal, QC, Canada, [3]Dept. of Neurology and Neurosurgery, McGill Univ., Montreal, QC, Canada

Disclosure Block: **P. Alipour:** None.

Background: selective vulnerability of some neuronal subpopulations is a common feature of neurodegenerative disorders. While upper and lower MNs degenerate in Amyotrophic Lateral sclerosis (ALS) ocular motor neurons including oculomotor nucleus (CNIII) seem to be largely spared. Diverse CNS cell types may contribute differently to neurodegeneration and are differentially vulnerable. Utilizing a single-nucleus RNA sequencing (snRNA-seq) method we aim to compare transcriptomic profile of differentially vulnerable regions to identify intrinsic properties of oculomotor nucleus that would be absent in medulla in ALS. Methods: Post-mortem CNS tissues from 5 ALS cases and 4 normal controls were obtained from the Douglas-Bell Canada Brain Bank, McGill university. Nuclei isolation was performed according to the Nagy et al. Following nuclei isolation, library preparations, sequencing, and Cellranger analyses were performed as per established protocol by 10X Genomics. Following the reception of the expression matrix, nuclei were analyzed in the single-cell analysis R package, Seurat. After quality control and clustering, the differentially gene expression was performed to compare oculomotor nucleus vs medulla in ALS patients. Results: clustering of 92K nuclei revealed 23 cell types. Endothelial cell and astrocytes are the only major cell types under-represented in the ALS samples. The majority of DEGs in oculomotor nucleus attribute to astrocytes and oligodendrocytes precursor cells (OPCs), while in the medulla the majority of DEGs attribute to neurons, astrocytes and OPCs. Glial cells show divergent signature in medulla vs oculomotor; while in oculomotor OPCs are involved in pathway indicating activation of the proteasome and phagosome system in response to proteinopathy in medulla OPCs are mostly enriched in antigen processing and presentation including major histocompatibility complex class I (MHC-I); microglia subpopulation signature in medulla represents Disease Associated Microglia (DAM) which is associated with reactive microglia. Conclusion: Based on the observation we made thus far, it appears that glial cells are active in the two ALS regions examined here but top DEGs identified in oculomotor do not reveal neurotoxic pathways but are rather associated with the maintenance of cell homeostasis. Given the neuronal dysregulation observed in oculomotor nucleus from ALS cases is minimum, by comparison to what is seen in medulla region, it is presumed that glial cells might be more supportive in this region and better deal with cellular stress of the disease.

PrgmNr 2848 - The influence of *APOE e*4 and global ancestry on self-reported anxiety and depression in Hispanic, Non-Hispanic Whites, and African Americans

View session detail

Author Block: **A. Zaman**[1], F. Rajabli[1], P. Mena[1], L. D. Adams[1], M. Contreras[1], F. Lacroix[1], S. Tejada[1], C. Silva-Vergara[2], M. Illanes-Manrique[3], J. Welch[1], M. R. Cornejo-Olivas[3], B. Feliciano[2], G. S. Byrd[4], J. L. Haines[5], J. M. Vance[1], M. A. Pericak-Vance[1], M. L. Cuccaro[1]; [1]Univ. of Miami, Miami, FL, [2]Univ. Central del Caribe, Bayamon, PR, [3]Inst. Natl. de Ciencias Neurologicas, Lima, Peru, [4]Maya Angelou Ctr. for Hlth.Equity, Winston-Salem, NC, [5]Case Western Reserve Univ., Cleveland, OH
Disclosure Block: **A. Zaman:** None.
Background: Anxiety (Ax) and Depression (Dp) co-occur with Alzheimer disease (AD) and are associated with a worse prognosis. There are limited data regarding the relationship between ancestry and race/ethnicity differences in Ax and Dp in AD. For this study we compared the prevalence of Ax and Dp and their association with global ancestry, self-reported race/ethnicity, *APOE e*4 presence, and phenotype [cognitively normal (CN), mild cognitive impairment (MCI), and AD]. Based on prior work we hypothesize that regardless of phenotype, Hispanics (HI), will have the highest prevalence of Ax and Dp followed by African Americans (AA) and Non-Hispanic Whites (NHW). Given that NHWs are predominantly of European (EU) ancestry we predict that EU ancestry will be negatively associated with Ax and Dp.
Methods: We identified 3597 older adults (mean age=74.2 Â± 9.0) from a study of AD genetics. Race/ethnicity were self-reported (HI=790; AA=481, NHW=2326). To estimate admixture proportion for each individual, we performed model-based clustering using ADMIXTURE software. *APOE* genotyping was performed as in Saunders (1996). Ax and Dp status (yes/no) were based on participant/family member report. Participants were classified as AD (n = 955), MCI (n = 514), or CN (n = 2128) by consensus diagnosis. A logistic regression examined Ax and Dp prevalence across race/ethnic groups (HI, AA, NHW), global ancestry [EU, African (AF), American Indian (AI)], *APOE e*4 presence, sex, age, and phenotype (AD, MCI, CN).
Results: Compared to NHW, HI had an increased prevalence of both Ax (p p 1) compared to EU ancestry, AF ancestry was associated with a lower prevalence of Ax (p = 0.001) and Dp (p = 0.015), while AI ancestry was only associated a lower prevalence of Ax (p 2) age was negatively associated with the prevalence of Ax (p p 3) compared to CN individuals, those with AD and MCI had a higher prevalence of Ax and Dp and females had a higher prevalence of Ax and Dp. *APOE e*4 was not associated with Ax or Dp.
Conclusion: The prevalence of Ax and Dp was greater in AA and HI compared to NHW while, at the same time, higher proportions of EU ancestry were associated with a greater prevalence of Ax and Dp. These novel results highlight the different contributions of self-reported race/ethnicity (e.g., sociocultural influences) and ancestry (e.g., genetic background) to neuropsychiatric phenotypes in AD. Moving forward, we need to explore how these contributions can inform the evaluation and treatment of Ax and Dp in underrepresented populations with AD to reduce the burden of these disorders.

PrgmNr 2849 - Two Genetic Variants Explain the Association of European Ancestry with Multiple Sclerosis Risk in African-Americans

View session detail

Author Block: **N. J. Nakatsuka**[1], N. Patterson[2], N. A. Patsopoulos[3], N. Altemose[4], A. Tandon[1], A. H. Beecham[5], J. McCauley[6], N. Isobe[7], S. Hauser[8], P. L. De Jager[9], D. Hafler[10], J. R. Oksenberg[11], D. E. Reich[1]; [1]Harvard Med. Sch., Boston, MA, [2]Harvard Univ., Cambridge, MA, [3]Brigham & Women's Hosp, Boston, MA, [4]UC Berkeley, Berkeley, CA, [5]Univ Miami, Miami, FL, [6]Univ. of Miami, Miami, FL, [7]UCSF, San Francisco, CA, [8]Univ. of California San Francisco Sch. of Med., San Francisco, CA, [9]Columbia Univ Med Ctr, New York, NY, [10]Yale Sch. of Med., New Haven, CT, [11]Univ California San Francisco, San Francisco, CA

Disclosure Block: **N.J. Nakatsuka:** None.

Epidemiological studies have suggested differences in the rate of multiple sclerosis (MS) in individuals of European ancestry compared to African ancestry, motivating genetic scans to identify genetic factors that could contribute to such patterns. In a whole-genome scan in 899 African-American cases and 1,155 African-American controls, we confirm that African-Americans who inherit segments of the genome of European ancestry at a chromosome 1 locus are at increased risk for MS [logarithm of odds (LOD) = 9.8], although the signal weakens when adding an additional 406 cases, reflecting heterogeneity in the two sets of cases [logarithm of odds (LOD) = 2.7]. The association in the 899 individuals can be fully explained by two variants previously associated with MS in European ancestry individuals. These variants tag a MS susceptibility haplotype associated with decreased *CD58* gene expression (odds ratio of 1.37; frequency of 84% in Europeans and 22% in West Africans for the tagging variant) as well as another haplotype near the *FCRL3* gene (odds ratio of 1.07; frequency of 49% in Europeans and 8% in West Africans). Controlling for all other genetic and environmental factors, the two variants predict a 1.44-fold higher rate of MS in European-Americans compared to African-Americans.@font-face {font-family:'Cambria Math'; panose-1:2 4 5 3 5 4 6 3 2 4; mso-font-charset:0; mso-generic-font-family:roman; mso-font-pitch:variable; mso-font-signature:-536870145 1107305727 0 0 415 0;}p.MsoNormal, li.MsoNormal, div.MsoNormal {mso-style-unhide:no; mso-style-qformat:yes; mso-style-parent:''; margin:0in; margin-bottom:.0001pt; mso-pagination:widow-orphan; font-size:12.0pt; font-family:'Times New Roman',serif; mso-fareast-font-family:'Times New Roman';}.MsoChpDefault {mso-style-type:export-only; mso-default-props:yes; font-family:'Cambria',serif; mso-ascii-font-family:Cambria; mso-ascii-theme-font:minor-latin; mso-fareast-font-family:'MS Mincho'; mso-fareast-theme-font:minor-fareast; mso-hansi-font-family:Cambria; mso-hansi-theme-font:minor-latin; mso-bidi-font-family:'Times New Roman'; mso-bidi-theme-font:minor-bidi;}div.WordSection1 {page:WordSection1;}

PrgmNr 2850 - Two-trait integrative analysis of de novo mutations in autism spectrum disorder and developmental disorders identifies shared CNS processes and epigenomic functions

View session detail

Author Block: **M. Kealhofer**[1], B. P. Riley[2], T-H. Nguyen[3]; [1]Virginia Commonwealth Univ. Sch. of Med., Richmond, VA, [2]Virginia Commonwealth Univ, Richmond, VA, [3]Virginia Commonwealth Univ., Richmond, VA

Disclosure Block: **M. Kealhofer:** None.

Exome studies of autism spectrum disorder (ASD) and other developmental disorders (DD) have implicated *de novo* mutations (DNMs) in overlapping genes and pathways. ASD and DD also share comorbidities, including intellectual disability and childhood onset seizures. In this analysis, we aim to use a two-trait analytic approach to increase statistical power for gene discovery based on the shared genetic etiology of ASD and DD. We jointly analyzed exome variants identified from 6,430 ASD trios and 31,058 trios with severe DD using a Bayesian integrative analysis method, mTADA (multi-trait transmission and de novo association test). We identified 164 ASD risk genes and 302 DD risk genes using a cutoff of 80% on their respective posterior probabilities. Among the ASD risk genes identified, 93 had not previously been implicated in single-trait analyses of this sample, although some were implicated in other ASD studies. Similarly, among the DD risk genes identified, 79 had not been identified in previous single-trait analyses. Among these genes, 153 were identified as being risk genes for both ASD and DD using the same posterior probability threshold. KEGG pathway analysis of these shared risk genes identified 20 enriched pathways after multiple test correction, including overlapping pathways for dopaminergic (p

PrgmNr 2852 - Assortative mating drives pathogenicity of variably expressive CNVs

View session detail

Author Block: **C. Smolen**[1], M. C. Jensen[1], L. Pizzo[1], A. Tyryshkina[1], V. Pounraja[1], L. Dyer[2], J. S. Juusola[2], 16p12.1 Consortium, S. Giririjan[1]; [1]Pennsylvania State Univ., University Park, PA, [2]GeneDx, Gaithersburg, MD
Disclosure Block: **C. Smolen:** None.
Many disease-associated copy-number variants (CNVs) are variably expressive, conferring risk for a range of neurodevelopmental and psychiatric features. For example, the developmental delay-associated 16p12.1 deletion is inherited in >90% of cases from a parent with mild features. We previously found that affected children were more likely to carry rare variants elsewhere in the genome compared to parents with the deletion, suggesting that the phenotypic variability observed in deletion carriers is driven by combinatorial effects with variants in the genetic background. We performed detailed phenotyping and whole genome sequencing of 125 families with the deletion to identify factors contributing to variable expressivity. The burden of secondary variants in probands was strongly associated with family history of neuropsychiatric phenotypes, likely due to increased transmission from non-carrier parents. We hypothesized that this increase in genetic burden and phenotypic severity may be driven by assortative mating, a form of non-random mating among individuals with similar genotypes or phenotypes. We found that 41/59 carrier and 29/65 non-carrier parents manifested neuropsychiatric features, and identified significant correlations between specific traits in pairs of parents. For example, individuals with depression or anxiety were more likely to pair with other individuals with similar phenotypes (OR=11, FDR=0.041). To further examine the role of assortative mating in driving phenotypic heterogeneity, we assessed genetic similarity (kinship coefficients) of 16p12.1 parents, and compared the results to parents from the Simons Simplex Collection (SSC) and Searchlight cohorts. Parental pairs in 16p12.1 cohorts had significantly higher kinship coefficients than random pairs within the cohort, and the degree of assortative mating was about three times higher compared to the SSC cohort (16p12.1 AUC: 0.141, SSC AUC: 0.051). This increase in kinship was not seen in families with autism-associated 16p11.2 CNVs, which mostly occur de novo. We further assessed parental kinship values in a larger set of 367 affected individuals carrying 21 disease-associated CNVs, and found higher kinship values for families with variably-expressive CNVs, such as 1q21.1 and 16p13.11 CNVs, compared to syndromic CNVs such as Smith-Magenis syndrome (p=0.033). Kinship values were also higher for families with inherited CNVs than de novo CNVs (p=0.014). These results suggest that assortative mating contributes to interactions between variably-expressive CNVs and additional variants in the genetic background to modulate the ultimate phenotypic trajectories.

PrgmNr 2853 - Clinical value of adding copy number, tandem repeat and antipsychotics metabolizer status analysis to diagnostic sequencing of patients with psychotic disorders

View session detail

Author Block: **A. Alkelai**[1], L. Greenbaum[2], G. Povysil[3], A. Malakar[3], S. Delaney[4], A. Grossman-Jonish[2], V. Jobanputra[5,6], A. Pulver[7], B. Lerer[8], D. B. Goldstein[9]; [1]Inst. for Genomic Med., Columbia Univ. Med. Ctr., New York, NY, [2]The Danek Gertner Inst. of Human Genetics, Sheba Med. Ctr., Tel Hashomer, Israel, [3]Inst. for Genomic Med., Columbia Univ., New York, NY, [4]New York State Psychiatric Inst., Columbia Univ., New York, NY, [5]New York Genome Ctr., New York, NY, [6]Dept. of Pathology and Cell Biology, Columbia Univ., New York, NY, [7]Johns Hopkins Sch Med., Baltimore, MD, [8]Biological Psychiatry Lab., Dept. of Psychiatry, Hadassahâ□□Hebrew Univ. Med. Ctr., Jerusalem, Israel, [9]Columbia Univ. Med. Ctr., New York, NY

Disclosure Block: **A. Alkelai:** None.

Background: Schizophrenia has a multifactorial etiology, involving a polygenic architecture. The potential yield of whole genome sequencing (WGS) in schizophrenia and other psychotic disorders is not well studied. Moreover, the benefit of different types of genome analyses is not systematically investigated. **Methods:** We investigated the yield of WGS analysis of single nucleotide variants (SNVs) or small insertions or deletions (indels) in 251 families with a proband diagnosed with psychotic disorder, by systematically applying the American College of Medical Genetics and Genomics (ACMG) criteria for categorization of variants. In addition, we used WGS data to analyze copy number variants (CNVs) and tandem repeats and determined antipsychotic metabolizer status by *CYP2D6* genotype data according to recommendations from the clinical pharmacogenetics implementation consortium. **Results:** The total diagnostic rate was 6.4%. We found clinically significant single nucleotide variants or indels in 14 families (5.6%) and CNVs in 2 families (0.8%). We did not detect any significant tandem repeats in patients with psychotic disorders. In addition, we were able to provide *CYP2D6* genotypes for most participants and determine their antipsychotic metabolizer status. 21.9% of our probands were found to be ultrarapid, intermediate or poor metabolizers, thus could potentially benefit from *CYP2D6* genotyping. **Discussion:** Taken together, the benefit of performing copy number and tandem repeats analyses using WGS data of patients diagnosed with psychotic disorders is very low. However, patients can potentially benefit from *CYP2D6* genotyping analysis. Yet, in order to provide actionable recommendations based on this data, integration of clinical information is required.

PrgmNr 2854 - Contribution of copy number variants to schizophrenia in East Asian populations

View session detail

Author Block: **Q. Feng**[1,2], B. Thiruvahindrapuram[3], D. P. Howrigan[4], Stanley Global Asia Initiatives, J. Sebat[5], T. Ge[6], H. Huang[4]; [1]Stanley Ctr. for Psychiatric Res., Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Dept. of Psychiatry, Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, Boston, MA, [3]The Ctr. for Applied Genomics, The Hosp. for Sick Children, Peter Gilgan Ctr. for Res. and Learning, Toronto, Toronto, ON, Canada, [4]Boston, MA, [5]UC San Diego, La Jolla, CA, [6]Harvard Med. Sch., Malden, MA

Disclosure Block: **Q. Feng:** None.

Schizophrenia is a chronic and severe mental disorder affecting approximately 1% of worldwide populations, with East Asians contributing the largest patient number. Recent studies have established an important role for copy number variants (CNVs) in the etiology of schizophrenia using individuals of European descent. However, to date, the extent to which CNVs contribute to schizophrenia in other continental ancestries is unclear. Here, we report findings from the largest CNV study outside of European ancestry, using subjects of East Asian ancestries from Japan, Korea, mainland China, Taiwan, Singapore and Indonesia. Stage 1 of this analyses included 12,056 schizophrenia cases and 12,978 controls. After extensive quality control harmonising datasets across various arrays, we observed a significant enrichment of genome-wide rare CNV burden (length >500kb, frequency

PrgmNr 2855 - Dissecting Biological Pathways of Psychopathology using Cognitive Genomics

View session detail

Author Block: **M. Lam**[1], C-Y. Chen[2], R. Tian[3], H. Huang[4], H. Runz[5], T. Ge[6], T. Lencz[7]; [1]The Broad Inst., Cambridge, MA, [2]Biogen, Cambridge, MA, [3]Malden, MA, [4]Boston, MA, [5]Biogen Inc., Cambridge, MA, [6]Harvard Med. Sch., Malden, MA, [7]Zucker Hillside Hosp., Glen Oaks, NY
Disclosure Block: **M. Lam:** None.
Cognitive deficits are known to be related to most forms of psychopathology. Recent work in the psychiatric and cognitive genomics literature indicated that pleiotropic associations between cognitive dimensions and psychiatric traits could further illuminate putative underlying biological mechanisms. Here, we perform local genetic correlations as means of identifying independent segments of the genome that might show biologically interpretable pleiotropic associations between cognitive dimensions and psychopathology. We utilized GWAS summary statistics for two major cognitive dimensions: *General Cognitive Ability (GCA)* and *Non-Cognitive Skills (*NCS*)*; the former indexes â□□*g*â□□ while the latter is derived from Educational Attainment after removing variance related to â□□*g*â□□. Applying local genetic correlation analysis, we were able to identify collective segments of the genome, defined as â□□meta-loci,â□□ that showed differential pleiotropic patterns for GCA and NCS against psychopathological. These meta-loci offer higher resolution and greater interpretability of the shared genetic architecture between cognitive dimensions and psychopathology as compared to global genetic correlations. Functional annotation and gene set analyses revealed a broad distinction between meta-loci associating psychopathology to GCA as compared to NCS, extending our prior work in schizophrenia: neurodevelopmental pathways predominated in GCA-relevant meta-loci, while synaptic pathways were more involved in NCS-relevant meta-loci. Driver genes related to GCA were found to be expressed during the prenatal-early childhood phase whereas those implicated with NCS were expressed in later childhood, young adulthood and beyond. Most notably, we found that many genes whose protein products were identified as presently druggable for psychiatric and nootropic indications were found within NCS meta-loci. These genes were predominantly within the GABA-ergic receptor, glutamate, and cholinergic family of genes.

PrgmNr 2856 - Enrichment of the VNTR *PER3* variant in DSWPD patients: A large whole genome sequencing analysis

View session detail

Author Block: **J. Brzezynski**, A. R. Kaden, J. A. Shinn, J. Wang, C. Xiao, S. P. Smieszek, C. Polymeropoulos, G. Birznieks, M. H. Polymeropoulos; Vanda Pharmaceuticals Inc., Washington, DC
Disclosure Block: **J. Brzezynski:** Salary/Employment; Vanda Pharmaceuticals Inc..
We have conducted a rigorous observational research study in suspected delayed sleep-wake phase disorder (DSWPD) patients. The objective was to measure sleep-wake patterns and to conduct exploratory genetic analyses to delineate the genetic landscape of the DSWPD phenotype.
We measured the sleep-wake patterns (self-reported bed time, wake time, midpoint of sleep, and sleep latency) of participants by daily post-sleep diaries for 10 weeks. Participants completed questionnaires on demographics, medical and surgical history, sleep history, and concomitant medications. Altogether, 119 participants were consented and 76 participants provided samples for whole genome sequencing. Seventy-eight participants were females and the mean age was 44 years. Principal component analysis defined ancestry as 29.3% AFR, 17.3% AMR, and 53.4% EUR.
We observed a significant enrichment of the minisatellite 54bp (1: 7829913-7829966 (GRCh38)) variable number of tandem repeat (VNTR) *PER3* rs57875989 4 allele. We observed significantly higher frequencies of the 4/4 variant, 59.2% when compared to the super control population frequency of 42.2% (n = 315; recessive: OR 1.9, CI 1.07 to 3.65, p This variant is of particular interest as it is located in the coding region of VNTR (exon 18). Functionally, *PER3* is phosphorylated by casein kinase 1 (CK1) and translocates to the nucleus to inhibit CLOCK/BMAL1 in the presence of *PER1*. The VNTR motif contains clusters of potential phosphorylation sites for CK1. Given the VNTR *PER3* variant could change protein phosphorylation levels in addition to tertiary protein structure and also have interactions with binding partners, it is hypothesized that the VNTR would cause functional changes in *PER3*. The observed accumulation of the *PER3* VNTR 4/4 supports prior literature describing evening-types and DSWPD patients as having greater frequencies of $PER3^{4/4}$ homozygotes; this effect can be as high as 75% homozygotes in DSWPD.
These results demonstrate that, on average, DSWPD individuals are more likely to harbor variants within their core clock genes with particular enrichment of the VNTR variant, potentially leading to a pronounced delay in sleep period. This variant can further impact the response to treatment across carriers of the 4/4.

PrgmNr 2857 - Extracellular vesicles: A window into the etiology of major depressive disorder

View session detail

Author Block: **P. Ibrahim**[1], P. Danthi[2], J-F. ThÃ©roux[2], C. E. Nagy[2], G. Turecki[3]; [1]IPN, McGill Univ., Douglas Mental Hlth.Univ. Inst., Montreal, QC, Canada, [2]Douglas Mental Hlth.Univ. Inst., Verdun, QC, Canada, [3]McGill Univ, Montreal, QC, Canada

Disclosure Block: **P. Ibrahim:** None.

Major Depressive Disorder (MDD) is one of the leading causes of disability worldwide, affecting 20% of the population. MDD also confers a 20-fold higher risk of suicide. Environmental factors are thought to play a role in disease development, resulting in biological changes mediated by epigenetic mechanisms. MicroRNAâ□□s (miRNA) are well known epigenetic regulators that are disrupted in the depressed brain, and they are packaged into extracellular vesicles (EVs). EVs have emerged as means of intercellular communication, a process that is also disrupted in MDD. They are thought to transfer miRNA, as well as other bioactive molecules such as proteins, between cells. This can alter gene expression in recipient cells. Different cell types in the brain have been shown to release EVs, so it is possible that EVs might play a role in the pathogenesis of central nervous system disorders, including MDD. Therefore, we hypothesize that EV cargo from the anterior cingulate cortex, a brain region highly implicated in MDD, will have a disease specific profile that could mediate disease development in MDD subjects compared to healthy controls (HC). Our objective is to isolate EVs from human post-mortem anterior cingulate cortex, profile the miRNA and protein cargo, and compare it between MDD subjects and HC. EVs were isolated from post-mortem human brain tissue from the anterior cingulate cortex of 43 MDD subjects and 43 HC using size exclusion chromatography. The quality was assessed by western blots and transmission electron microscopy (TEM). RNA was extracted and a small-RNA library was constructed and sequenced using the Illumina Platform. Proteins were also extracted and profiled using LC-MS/MS. Differential expression analysis was then performed. Western blots showed little to no contamination with cellular debris, along with enrichment of the exosomal marker CD9. TEM images showed the typical cup-shaped morphology with sizes mostly between 30 and 200 nm, and vesicles were labelled with the exosomal marker CD81. Preliminary differential analyses revealed sex-dependent dysregulation in miR-4485-3p, miR-142-5p, miR-33a-5p, miR-132-5p, and miR-92a-3p, as well as in proteomic profiles of the EVs in MDD. In conclusion, this will be the first study to profile brain-derived EV miRNA and protein in the context of depression. Future studies will be needed to determine the effect of the dysregulated EV cargo in MDD. This could provide novel mechanistic insights into the pathophysiology of MDD, which could serve as a starting point for the development of targeted therapeutic strategies as well as prevention measures.

PrgmNr 2858 - Genome-wide genetic susceptibility to pain and depression predicts suicidal thoughts and attempts inschool-agedchildren

View session detail

Author Block: **P. H. Lee**[1,2], A. E. Doyle[1,2], X. Li[1], M. D. Silberstein[1], J-Y. Jung[3], A. Nierenberg[1,2], R. Liu[1,2], R. Kessler[2], R. H. Perlis[1,2], M. Fava[1,2]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Stanford Univ., Stanford, CA

Disclosure Block: **P.H. Lee:** None.

Background: Suicide is among the leading cause of death in children and adolescents. There are well-known risk factors of suicide, including abuse, family conflicts, social adversity, and psychopathology. While suicide risk is also known to be heritable, few studies have investigated genetic risk in younger individuals.

Methods: Using polygenic risk score (PRS) analysis, we examined whether genetic susceptibility to chronic pain and major depression, two clinical conditions that have been linked to increased suicide risk, is associated with suicidal thoughts and behaviors (STBs) among 11,878 children enrolled in the Adolescent Brain Cognitive Development study (ABCD). The ABCD study represents school-aged children of diverse demographic and socioeconomic backgrounds in the USA, for which a wide range of social, familial, physical, mental, and behavioral aspects have been collected longitudinally since 2016, along with genome-wide genotype data. STBs were assessed using the Kiddie Schedule for Affective Disorder and Schizophrenia (KSADS-5). After performing robust quality control of genotype data, unrelated individuals of European descent were included in analyses (N=4,405). The average age of the children used in the analyses was 9.91 (0.62) years old in the baseline (47.84% female).

Results: Depression PRSs were associated with lifetime suicide attempts reported by children (*OR*=1.49, 95% *CI*=1.11-2.00, *p*-value=$8.65x10^{-3}$), after adjusting for childrenâ□□s sociodemographic backgrounds, family history of suicide, and psychopathology measured using the Child Behavior Checklist (CBCL). The *OR* represents the odds of inclusion in the suicidal group with an increase of one standard deviation change in the PRS. In contrast, pain PRSs were associated with lifetime suicidal ideation (*OR*=1.29, 95% *CI*=1.10-1.51, *p*-value=$9.36x10^{-4}$), suggesting a distinct contribution of the genetic risk underlying pain and depression on suicidal behaviors of children. Association of pain susceptibility was particularly strong for active suicidal thoughts involving specific methods, which remain significant after accounting for childrenâ□□s somatic health complaints.

Conclusion: The largest genetic sample of suicidality data in US children suggests a significant genetic basis of suicidality related to pain and depression. Although the predictive utility of suicide risk factors, including pain and depression PRSs, is still modest, future efforts may benefit from utilizing genetic risk factors when developing proactive screening and early intervention strategies for mitigating suicide risk in children.

PrgmNr 2859 - Identifying major depressive disorder subtypes using polygenic risk scores

View session detail

Author Block: **C. Zai**[1], A. K. Tiwari[1], G. C. Zai[1], D. J. Mueller[1], N. Freeman[1], N. King[1], S. Y. Cheema[1], D. Herbert[1], H. Emmerson[1], J. L. Kennedy[2]; [1]Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [2]Ctr Addiction & Mental Hlth, Toronto, ON, Canada

Disclosure Block: **C. Zai:** Receipt of Intellectual Property Rights/Patent Holder; patents and patent applications for antipsychotic-induced weight gain and suicide markers. Other; Attendance at American Society of Clinical Psychopharmacology annual meeting 2021 supported by Myriad Neuroscience.

Major depressive disorder (MDD) is a serious psychiatric disorder and a leading cause of disability worldwide. Its etiopathophysiology is complex and not well understood. While MDD has a genetic component, it is likely highly polygenic. In order to better understand the complexities surrounding MDD, we carried out a polygenic risk score analysis of 1,171 MDD patients from the Individualized Medicine: Pharmacogenetic Assessment & Clinical Treatment (IMPACT) study. We performed K-means clustering with cluster number ranging from two to five using polygenic risk scores for MDD, bipolar disorder, educational attainment, insomnia, alcohol addiction, maltreatment, impulsivity, loneliness, and post-traumatic stress disorder (JMP v15). We found the two-cluster solution to have the highest Cubic Clustering Criterion. One cluster has higher average polygenic risk scores for impulsivity, alcohol addiction, and loneliness, and lower average scores for MDD, bipolar disorder, educational attainment, and insomnia than the other cluster (p

PrgmNr 2861 - Pathways Implicated in Risk for Suicide Attempts in the Million Veterans Program

View session detail

Author Block: **E. R. Hauser**[1,2], X. J. Qin[3], A. E. Ashley-Koch[4], M. A. Hauser[2], B. H. McMahon[5], J. H. Lindquist[1], R. Madduri[6], J. E. Huffman[7], N. A. Kimbrel[1,2], J. C. Beckham[1,2], and the MVP Suicide Exemplar Workgroup; [1]Durham VA Hlth.Care System, Durham, NC, [2]Duke Univ., Durham, NC, [3]Durham VA Hlth.Care System, Durham, NC, [4]Duke Univ Med Ctr, Durham, NC, [5]Los Alamos Natl. Lab., Los Alamos, NM, [6]Argonne Natl. Lab., Lemont, IL, [7]VA Boston Hlth.care System, Jamaica Plain, MA
Disclosure Block: **E.R. Hauser:** None.
We conducted a genome-wide association study (GWAS) of suicide attempts in the large, multi-ethnic Million Veterans Program (MVP), comparing U.S. veterans with a documented history of suicide attempts (N=14,089) to U.S. veterans with no documented history of suicidal thoughts or behaviors (N=395,359). Analyses were conducted within four populations (European, African American, Asian, and Hispanic) as well as a multi-ethnic meta-analysis. To identify specific biological pathways implicated in risk for suicide attempts, we applied Over-Representation Analysis (Boyle et al. 2004) in the Web-Gestalt package, defining pathways with the KEGG database. Within the multi-ethnic meta-analysis, multiple pathways converged on mood alteration and stress. Most significantly over-represented was oxytocin signaling (FDR $P=1.9x10^{-5}$), which plays a key role in social bonding and feelings of well-beingâ□□disruptions of this pathway could lead to feelings of despair and suicidal behavior. Multiple stress pathways were represented, including cortisol synthesis and secretion, as well as blood pressure regulating pathways (renin secretion, aldosterone synthesis and secretion, and vascular smooth multiple contraction). There is striking consistency across analyses of individual racial groups, with more than half of the 30 pathways implicated in the meta-analysis also significantly over-represented in both Caucasians and African Americans. Pathways significant in all 5 analyses include arrhythmogenic right ventricular cardiomyopathy (ARVC), and phospholipase D signaling. Among the most interesting pathways implicated in individuals with a history of suicide attempts were the circadian entrainment and circadian rhythm pathways. Disruptions of circadian rhythm profoundly affects mood, as CLOCK protein regulates dopaminergic transmission in the ventral tegmental areaâ□□the critical neurological reward circuits. Based on the over-representation of circadian rhythm pathways in the GWAS data, we examined the prevalence of sleep disorders within the MVP GWAS cohort, as defined by the VITAL (VITamins And Lifestyle) questionnaire. Individuals with a history of suicide attempts reported 3.5 (S.D.=2.2) sleep problems, compared with 2.2 (S.D. 2.0) in controls, strongly corroborating the pathway analysis. In summary, while ethnic-specific GWAS identified different genetic variants and loci, pathway analysis across all populations are in broad agreement on the biological factors underlying increased risk of suicide attempts.

PrgmNr 2862 - Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia, Bipolar Disorder, and Major Depression in 660,000 US Veterans

View session detail

Author Block: **T. B. Bigdeli**[1,2], G. Voloudakis[3,4], P. B. Barr[5], Y. Li[6], N. Rajeevan[7], F. Sayward[7], B. Gorman[8], T. Oâ□□Leary[9], T. Gleason[10], R. Przygodzki[11], G. D. Huang[9], S. Pyarajan[12], S. Muralidhar[13], J. M. Gaziano[14], Cooperative Studies Program #572, Million Veteran Program (MVP), M. ASLAN[7], A. Fanous[15], P. D. Harvey[16], P. Roussos[3,4]; [1]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [2]VA NY Harbor Hlth.care System, Brooklyn, NY, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]James J. Peters VA Med. Ctr., Bronx, NY, [5]Virginia Commonwealth Univ., Richmond, VA, [6]VA Clinical Epidemiology Res. Ctr. (CERC), VA Connecticut Hlthcare System,, West Haven, CT, [7]VA Clinical Epidemiology Res. Ctr. (CERC), VA Connecticut Hlthcare System, West Haven, CT, [8]VA Boston Hlth.care System, Boston, MA, [9]Office of Res. and Dev., Veterans Hlth.Admin., Washington, DC, [10]Clinical Sci. Res. and Dev. Service, Office of Res. and Dev., VA Central Office, Washington, DC, [11]Washington, [12]VA/HMS, Lexington, MA, [13]Dept VA, Washington, DC, [14]VA Boston Hlth.care System, Roxbury Crossing, MA, [15]SUNY Downstate Hlth.Sci. Univ. & Brooklyn VA, Brooklyn, NY, [16]Miami VA Hlth.care System, Miami, FL

Disclosure Block: **T.B. Bigdeli:** None.

Background/Aims: Serious mental illness, including schizophrenia, bipolar disorder, and major depression are heritable, highly multifactorial disorders and major causes of disability worldwide. Polygenic risk scores (PRS) aggregate variants identified from genome-wide association studies (GWAS) into individual-level estimates of liability to disease, and are a promising tool for risk stratification in clinical settings.

Methods: Leveraging the VAâ□□s extensive electronic health record (EHR) and an embedded cohort of 9,378 individuals with confirmed diagnoses of schizophrenia or bipolar I disorder, we validated automated case-control assignments based on ICD-9/10 codes, and benchmarked the performance of current PRS in 660,000 Million Veteran Program (MVP) participants. We explored broader relationships between PRS and 1,700 disease categories and 70 laboratory values via phenome-wide association studies (PheWAS). Given the substantial genetic overlap between neuropsychiatric disorders, we applied genomic structural equation modeling (gSEM) to derive novel PRS indexing common and disorder-specific latent genetic factors.

Results: Among 3,953 and 5,425 individuals with confirmed diagnoses of schizophrenia or bipolar disorder, 95% were correctly identified using ICD-9/10 codes (2 or more), with 25% also meeting criteria for the incorrect diagnosis. Encouragingly, PRS performed best in comparisons of confirmed cases, followed by individuals who had received inpatient treatment, individuals who received any treatment, and persons with spectrum diagnoses. Our PheWAS confirmed previous reports that individuals with higher neuropsychiatric PRS are at elevated risk for a range of psychiatric and physical health problems, and we demonstrate that many of these findings are generalizable to African Americans. In our novel LabWAS, we observed robust, positive associations between schizophrenia PRS and cholesterol levels, white blood cell count, and glomerular filtration rate, and significant negative associations with serum glucose and electrolyte levels.

Conclusions: Using current neuropsychiatric PRS, we have demonstrated the validity of EHR- based phenotyping approaches across diverse US populations. Our PheWAS uncovered novel associations and replicated previously reported relationships between schizophrenia, bipolar disorder, and major depression PRS and a range of clinical traits and outcomes, highlighting the potential of PRS for disentangling biological and mediated pleiotropy.

PrgmNr 2863 - Perspectives on the COVID-19 pandemic: Report from individuals with copy number variations and their caregivers

View session detail

Author Block: **L. White**[1], T. Crowley[2], B. M. Finucane[3], S. Garcia-Minaur[4], G. M. Repetto[5], M. Fischer[6], M. van den Bree[7], S. Jacquemont[8], R. Barzilay[1], A. M. Maillard[9], K. A. Donald[10], R. E. Gur[1], A. S. Bassett[11], A. Swillen[12], D. M. McDonald-McGinn[13]; [1]Children's Hosp. of Philadelphia and Perelman Sch. of Med. of the Univ. of Pennsylvani, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Geisinger Hlth.System, Lewisburg, PA, [4]Hosp. Univ.rio La Paz, Madrid, Spain, [5]Clin Alemana- Univ Desarrollo, Santiago, Chile, [6]Univ. of Rostock, Rostock, Germany, [7]Cardiff Univ., Cardiff, United Kingdom, [8]CHU Sainte justine Res. Ctr., Montreal, QC, Canada, [9]Service des Troubles du Spectre de lâ□□Autisme (STSA), Lausanne Univ. Hosp., Lausanne, Switzerland, [10]Red Cross War Mem. Childrenâ□□s Hosp., Univ. of Cape Town, Cape Town, South Africa, [11]Univ of Toronto, Toronto, ON, Canada, [12]KU Leuven, Leuven, Belgium, [13]Children S Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **L. White:** None.
Background: The world has suffered immeasurably during COVID-19, with concomitant behavioral health problems. University of Pennsylvania (UPenn) investigators reported a 1SD increase in COVID-19 worries associated with a 2-fold increase in generalized anxiety, 67% increase in depression, and an escalation of postpartum depression. The G2MH Network sought to understand how stakeholders affected by copy number variations (CNVs), including 22q11.2, were coping during the pandemic. **Methods:** The UPenn abbreviated COVID-19 Distress Survey was circulated by 22 advocacy groups, including 22q11.2 organizations, in English and 11 other languages. **Results:** 663 people from 26 countries completed the survey. Most respondents were mothers of children with 22q11.2 CNVs. Top worries included: family members acquiring COVID, unknowingly transmitting/personally acquiring COVID, financial pandemic-related burdens, dying from COVID, and currently having COVID. Total COVID-19 distress was higher in individuals completing the survey towards the end of the study (later pandemic wave) and individuals living in urban environments. There were no differences by region, CNV type, respondent type, timing of the survey, sex or age. A significant minority of participants (37%, n=253) reported pandemic realted healthcare interruptions had significant negative effects on their health. A total of 46% (N=305) of participants receiving hospital-based primary care delayed appointments due to COVID-19 fears and over half of participants (n=341) worried about future treatment disruption if the pandemic continued. Medical care disruptions, fears, and avoidance were related to higher COVID-19 distress. **Conclusions:** Effects of the COVID-19 pandemic for rare CNVs were broadly consistent with research results from other patient populations. Long term effects of COVID-19 distress, interruptions to care, and hospital avoidance require further study.

PrgmNr 2864 - Suicide attempts among European ancestry veterans in the Million Veteran Program exhibits correlation with multiple psychiatric conditions

View session detail

Author Block: **X. Qin**; Durham VA Hlth.Care System, Durham, NC
Disclosure Block: **X. Qin:** None.
Background Nearly 800 000 people die due to suicide globally every year, and the rate of suicide among U.S. military veterans is now 1.5 times higher than the rate of suicide among their civilian counterparts. For each adult who died by suicide, there may have been more than 20 others who attempted suicide. There is growing evidence that familial and genetic factors contribute to the risk for suicidal behavior. Moreover, major psychiatric illnesses, including bipolar disorder, major depression, schizophrenia, alcoholism and substance abuse, as well as certain personality traits are associated with suicidal behavior. Calculation of the genetic correlation between suicide attempts (SA) and different phenotypes is an efficient way to understand the genetic pleiotropy that leads to the phenotypic correlations.
Methods: The MVP cohort included 9196 veterans who made a SA and 287577 controls, all of European ancestry (EA). GWAS of SA was conducted using plink, adjusting for age, gender and genetic principle components. Several GWASs in EA for psychiatric illnesses were publicly available in ldhub, Psychiatric Genomics Consortium, as well as NEALE LAB-UKbiobank. For the purpose of this analysis, we examined bipolar, major depressive disorder, PTSD, schizophrenia, insomnia, sleep disorder. The genetic correlation between SA GWAS and psychiatric illnesses GWAS was calculated using LD score regression (LDSC).
Results: Some phenotypes showed strong significant genetic correlation with SA. The genetic correlation with bipolar was 0.38 (p-value=4.5E-09), with MDD was 0.63 (p-value 1.9E-20), with PTSD was 0.57 (p=5.59E-15), with schizophrenia was 0.46 (p=9.23E-16), and with insomnia was 0.096(p=5.7E-03).
Conclusions /Discussion SA among veterans with European ancestry in the MVP cohort exhibits strong genetic correlation with several psychiatric phenotypes that are often comorbid with SA, including MDD, PTSD, schizophrenia and bipolar disease. There is also limited, but significant genetic correlation between SA and sleep disorders. These observations should guide future genetic studies to elucidate the underlying genetic etiology of SA.

PrgmNr 2865 - Whole exome sequencing meta-analysis of depression suggests subtle role for functional variants in known GWAS loci

View session detail

Author Block: **A. Moscati**, V. M. Rajagopal, S. A. Di Gioia, L. Dobbyn, G. Coppola, E. Stahl; Regeneron Genetics Ctr., Tarrytown, NY
Disclosure Block: **A. Moscati:** Salary/Employment; Regeneron Pharmaceuticals.
Major depression (MD) is one of the most prominent causes of morbidity in the modern world and perhaps the single greatest source of years lived with disability, responsible for 7.8% of all disability in the Americas in 2018. In addition to the profound distress and impairment for the individual with MD, the aggregate public health impact is substantial, with the total annual economic burden of major depression in recent years estimated at over $200 billion in the United States. MD is a multifactorial disorder, with well-established risk factors including stressful life events, chronic sleep disturbance, and substance use. There are also genetic influences on MD, as shown by family studies using biometrical models. The availability of large genetic datasets with expansive clinical information has enabled successful analyses of genetic factors in association with MD, but most of these have been conducted on imputed genome-wide array data predominantly in European populations, which may miss lower-frequency and functional variants, or variants in other populations that could contribute to the disorder. We generated a composite phenotype across multiple depression measures in the UK Biobank (UKB) sample, informed by the work of Howard et al. 2019, and Glanville et al. 2021. We also created depression phenotypes with the published eMERGE algorithm in two hospital biobank cohorts, Geisinger Health Systemâ□□s DiscovEHR, and Mount Sinai Hospitalâ□□s Bio*Me* Biobank. We then used these phenotypes across the three datasets to conduct a genetic association analysis in individuals of all ancestral backgrounds on whole exome sequenced and genome-wide imputed data. We find 68 genome-wide significant loci, 25 of which are not previously reported to the authorsâ□□ knowledge, with 7 of these having non-synonymous or splice region variants in strong linkage disequilibrium (LD) with the index variants, including within the genes *TYR* and *TCF4*, and in 7 genes within the major histocompatibility complex (MHC) region; while 12 additional regions have variants that underlie expression quantitative trait loci (eQTL). Notably, as previously reported for other psychiatric disorders such as schizophrenia, we find a strong signal in the MHC region on chromosome 6, a region known for its complex, strong LD patterns. To discern which variants may be most responsible for these associations in the MHC and each other locus genome-wide, we use FINEMAP to prioritize credible sets of potentially casual variants. Our results suggest that functional and low-frequency variants may contribute to MD in addition to common variants of small effect in known genetic association loci.

PrgmNr 2866 - Whole-exome sequencing in the UK Biobank reveals risk gene *SLC2A1* and biological insights for major depressive disorder

View session detail

Author Block: **R. Tian**[1], T. Ge[2], J. Z. Liu[3], M. Lam[4], Biogen Biobank team, D. F. Levey[5], J. Gelernter[5], M. B. Stein[6], E. A. Tsai[1], H. Huang[7], T. Lencz[8], H. Runz[9], C-Y. Chen[1]; [1]Biogen, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]GSK, Philadelphia, PA, [4]The Broad Inst., Cambridge, MA, [5]Yale Univ. Sch. of Med., New Haven, CT, [6]Univ. of California San Diego, La Jolla, CA, [7]Broad Inst. of MIT and Harvard, Cambridge, MA, [8]Zucker Hillside Hosp., Glen Oaks, NY, [9]Biogen Inc., Cambridge, MA
Disclosure Block: **R. Tian:** Salary/Employment; Biogen.
Major depressive disorder (MDD) is a common and polygenic psychiatric disorder. Nearly two hundred depression risk loci have been identified by genome-wide association studies (GWAS). However, the impact of rare coding variants on depression remains poorly understood. Here, we present the largest to date exome analysis of depression based on 320,356 UK Biobank participants. We showed that the burden of rare protein-truncating coding variants and damaging missense variants in loss-of-function intolerant genes (pLI>0.9) conferred risk to a range of depression phenotypes as defined by different degrees of diagnostic stringency, with the most prominent effect in electronic health record (EHR)-defined depression (OR=1.17, 95% CI=1.13-1.21, *P*=3.57e-18). Among all depression definitions, we also found the strongest association between damaging missense variant burden and EHR-defined depression risk (OR=1.08, 95% CI=1.05-1.23, *P*=8.52e-6). We performed gene-based protein-truncating and damaging missense variant burden tests across the exome in the EHR-defined cohort (Ncase=10,449; Ncontrol=246,719), and identified *SLC2A1* (*P*=2.96e-07), which exceeded exome-wide significance, and a total of 30 risk genes with false discovery rate (FDR) SLC2A1 (*GLUT1*) encodes glucose transporter and facilitates transport of glucose into the brain cross brain-blood barrier. Mutations of *SLC2A1* impair energy supply for the brain and cause GLUT1 syndrome, characterized by developmental delay. Using human brain expression profiles and 10,271 gene sets (biological process, cellular component and molecular function) from MSigDB v7.2, we replicated known biological pathways of neuron projection development and muscle activities identified in depression GWAS. By jointly analyzing polygenic scores with exome data, we demonstrated additive contributions of common and rare coding variants to depression risk. Finally, we observed elevated genetic risk for depression in genes implicated by exome analyses of developmental disorder, autism and schizophrenia. Our study provides novel insights into the genetic architecture and biology of depression, and the genetic relationships across developmental and psychiatric disorders.

PrgmNr 2867 - A genetic epidemiologic study of myalgic encephalomyelitis (ME)/chronic fatigue syndrome (CFS) identifies links with autoimmune disease and cancer

View session detail

Author Block: **R. Moslehi**[1], A. Kumar[2], A. Dzutsev[3]; [1]Univ. at Albany, State Univ. of New York (SUNY), Albany, NY, [2]Univ. at Albany, SUNY, Albany, NY, [3]Natl. Cancer Inst. (NCI), NIH (NIH), Bethesda, MD
Disclosure Block: **R. Moslehi:** None.
Background: Myalgic encephalomyelitis (ME)/chronic fatigue syndrome (CFS) is a disabling multi-system complex disorder with prevalence of 875 per 100,000 (up to 3.4 million people) in the United States. There are no known etiologic or risk factors and no approved treatments for ME/CFS. We conducted a genetic epidemiologic study in order to test the hypothesis that ME/CFS may be an autoimmune disease (AID), and to explore the link between ME/CFS and cancer, specifically hematologic cancers. Methods: Our clinic-based study involved carefully selected cases with confirmed diagnosis of ME/CFS (n=59) and healthy controls (n=54) frequency matched to cases on age, gender, and ethnicity. During structured interviews, detailed multigeneration pedigrees and answers to an epidemiologic questionnaire were obtained on all subjects. Data analysis involved comparison of cases and controls with respect to the prevalence and cumulative incidence of AID and cancer among their first-degree relatives by calculating risk ratios, 95% confidence intervals (CI) and p values. Odd ratios (OR) were calculated for comparison of epidemiologic factors. Results: The prevalence of AID was significantly higher among first-degree relatives of cases compared to those of controls (OR=5.30; 95%CI: 1.83-15.38; p=0.001). The prevalence of AID among mothers was 14% for cases and 1.9% for controls (p=0.03). 11.2% of first-degree relatives of cases had an AID compared to 3.1% of relatives of controls (prevalence ratio=3.71; 95% CI: 1.74-7.88; p=0.0007); age-adjusted analysis was also significant (p=0.0025). Similar patterns of increased risk among the relatives of cases were found for cancer in general and hematologic cancers specifically. 20% of first-degree relatives of cases had any type of cancer compared to 15.4% of the relatives of controls (p=0.03). The cumulative incidence of hematologic cancers was significantly higher among the relatives of cases (p

PrgmNr 2868 - A multi-model data approach to improve statistics for Metabolite-Variant-Disease (MVD) relationship in SOL dataset

View session detail

Author Block: **P. Sharma**[1,2], C. Bizon[1,2], E. Morris[2], K. Bradford[1,2], K. E. North[1], P. Gordon-Larsen[3], M. Graff[1], H. M. Highland[4], A. Howard[1]; [1]Univ North Carolina, Chapel Hill, NC, [2]Renaissance Computing Inst. at UNC, Chapel Hill, NC, [3]Univ North Carolina at Chapel Hil, Chapel Hill, NC, [4]Univ North Carolina at Chapel Hill, Chapel Hill, NC

Disclosure Block: **P. Sharma:** None.

Independent GWAS and MWAS studies have been carried out to better understand the association between phenotype, genetics and metabolites on various sample population with obesity disease, is a complex and heterogeneous disease. We used a novel multi-model data approach via a knowledge graph (ObesityHub) to study the associations of metabolites (M), variants (V), and obesity disease (D), and improve the statistics by using a large ancestrally diverse cohort of Hispanic Americans. The Hispanic Community Health Study / Study of Latinos (HCHS/SOL) is a multi-center epidemiologic study in Hispanic/Latino populations that was conducted between 2008-2011 with study participants aged between 18-74 years. The meta-analysis from GWAS-associations with phenotypes, MWAS-associations with phenotypes, and GWAS-associations with metabolite was carried out and loaded into the existing knowledge graph (KG) in NEO4J, which also has multiple existing knowledge sources e.g., GTex, KEGG. Conducting GWAS or MWAS-associations with phenotypes alone reveals the associations between phenotypes and obesity disease, however, does not provide information about the underlying biological pathway mechanisms. Thus, by applying MWAS- and GWAS-associations with phenotypes and metabolites for the SOL cohort, we leverage the associations between the M, V and D for the same population via a KG. In addition, we were able to leverage the existing knowledge sources to find associations that drive variants and chemical interactions that affect obesity within the SOL population. We performed 3-way Cypher queries between a M, V, D, which we identified as â□□triangleså□□ of associations, based on a threshold of p-value

PrgmNr 2869 - A reliable algorithm to reconstruct modifier haplotypes in family-based studies of complex disorders using variable ages-at-onset for primary open-angle glaucoma as a disease model

View session detail

Author Block: **V. Raymond**[1,2], P. Belleau[3], S. Dubois[1], S. Desjardins[1], R. Arsenault[1], E. Shink[1], J-L. Anctil[2], G. CÃ´tÃ©[2], M. Amyot[4], M. A. Walter[5]; [1]CHUL Res. Ctr., Quebec City, QC, Canada, [2]UniversitÃ© Laval, Quebec City, QC, Canada, [3]Cold Spring Harbor Lab., Cold Spring Harbor, NY, [4]UniversitÃ© de MontrÃ©al, Montreal, QC, Canada, [5]Univ Alberta, Edmonton, AB, Canada
Disclosure Block: **V. Raymond:** None.
Epistasis is a mechanism by which the effect of a disease-causing gene mutation is dependent on mutations in one or more other genes, called modifiers. Finding modifier genes is a major challenge in humans as there is a lack of efficient strategies to reach this goal. To counteract this impediment, we developed a novel unbiased pedigree-based strategy to study large French-Canadian families affected by primary open-angle glaucoma (POAG), a highly prevalent and genetically complex disorder of the eye. Using genome-wide linkage studies and deep phenotyping in 156 heterozygotes for the autosomal dominant glaucoma myocilin K423E mutation (*MYOC*^{K423E}), we previously mapped a modiï¬er locus for variability of ages-at-onset (AAO) at chromosome 20q13. The locus is named, here, *MOG1* for *Modifier of Glaucoma 1*. Towards the characterization of the *MOG1* gene, we designed an algorithm that optimizes the identiï¬cation of reliable double-mutants who simultaneously carry the *MYOC*^{K423E} disease mutation and one (1) *MOG1* marker haplotype, or one (1) of its alleles, which is associated with extreme values for AAO. The method is named DIGGI for Double-mutants that participate In Gene-Gene Interactions. DIGGI is a two-stage algorithm that exploits datasets obtained with diverse types of markers that can be coded like microsatellites, VNTRs, SNPs and CNVs. The 1st stage is the identiï¬cation of *MOG1* alleles, or *MOG1* marker haplotypes, associated with the modiï¬er. The 2nd stage is the identiï¬cation of double-mutants who simultaneously carry the primary disease mutation and show higher contrasting ages at onset when compared to the AAO of individuals who are within their neighborhood (kinship coefficient defined as Î¦(X,Y) â¥ 0.0625 meaning that X and Y are people closer or equal to first degree cousins) and carry the K423E myocilin mutation. From our pool of *MYOC*^{K423E} heterozygotes, we will keep for further analysis as double-mutants (i.e., sequencing) only those who are double-mutants at the extremes of the extremes contrast distribution of the AAOs within their neighborhood. Conversely, we will keep as controls, *MYOC*^{K423E} carriers who do not harbour *MOG1* alleles, or *MOG1* marker haplotypes, at these opposite extremes for AAO. Combined with simulation and statistical studies, our data show that this algorithm is reliable as it selects unequivocal double-mutants that participate in gene-gene interactions. In conclusion, we developed and successfully applied to the case of glaucoma modiï¬er gene a powerful strategy to identify family members who share common modifier haplotypes, or modifier alleles, associated with speciï¬c endophenotypes for quantitative traits.

PrgmNr 2870 - Association of Clonal Hematopoiesis with Chronic Obstructive Pulmonary Disease

View session detail

Author Block: **D. Qiao**[1], P. G. Miller[2,3,4], J. Rojas-Quintero[1], M. C. Honigberg[5,6], A. S. Sperling[2,7], C. J. Gilbson[2], A. Bick[8], A. Niroula[2,3], M. E. McConkey[2], B. Sandoval[2], G. T. O'Connor[9], S. Redline[10,11], S. S. Rich[12], D. S. Neuberg[13], S. Jaiswal[14], B. D. Levy[1], C. A. Owen[10], P. Natarajan[15,6], E. K. Silverman[16], P. v. Galen[10], Y. Tesfaigzi[10], B. L. Ebert[17,3,18], M. H. Cho[10], COPDGene Study Investigators, NHLBI Trans-Omics for Precision Medicine (TOPMed)Consortium; [1]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [2]Dana-Farber Cancer Inst., Boston, MA, [3]Broad Inst. of MIT and Harvard Univ., Cambridge, MA, [4]Brigham and Womenâ□□s Hosp. and Harvard Med. Sch., Boston, MA, [5]Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [6]Broad Inst. of Harvard, Cambridge, MA, [7]Brigham and Womenâ□□s Hosp., Boston, MA, [8]Vanderbilt Univ. Med. Ctr., Nashville, TN, [9]Boston Univ. Sch. of Med., Boston, MA, [10]Brigham and Women's Hosp., Boston, MA, [11]Beth Israel Deaconess Med. Ctr., Boston, MA, [12]Univ Virginia, Charlottesville, VA, [13]Dana Farber Cancer Inst, Boston, MA, [14]Stanford Univ., Palo Alto, CA, [15]Massachusetts Gen. Hosp., Boston, MA, [16]Brigham & Women's Hosp, Boston, MA, [17]Dana-Farber Cancer Inst.; Broad Inst. of MIT and Harvard Univ.; Howard Hughes Med. Inst., Boston, MA, [18]Howard Hughes Med. Inst., Bethesda, MA

Disclosure Block: **D. Qiao:** None.

Background: Chronic obstructive pulmonary disease (COPD) is associated with age and smoking, but other determinants of the disease are incompletely understood. Clonal hematopoiesis of indeterminate potential (CHIP) is a common, age-related state in which somatic mutations in clonal blood populations induce aberrant inflammatory responses. Patients with CHIP have an elevated risk for cardiovascular disease, but the association with COPD remains unclear. **Methods:** We analyzed whole-genome and exome sequencing data to detect the presence of CHIP in 48,835 subjects, of whom 8,444 had moderate-to-very-severe COPD. Cases of CHIP were defined based on somatic mutations detected using the sequencing data that are in a pre-specified list of variants predicted or reported to be pathogenic and drivers of myeloid malignancies. The subjects were selected from four separate cohorts with COPD phenotyping and smoking history, including COPDGene, additional TOPMed cohorts, ICGN-EOCOPD study, and UK Biobank. We applied logistic regression and generalized linear mixed effect models to COPD status adjusting for age, gender, sequencing center, number of pack-years, smoking status, and genetic ancestry. Random-effects meta-analyses were conducted to combine the cohorts. As *TET2* was commonly mutated in hematopoietic cells, we measured emphysema in murine models in which *Tet2* was deleted in hematopoietic cells and performed single cell RNA-sequencing in lung tissues of the Tet2-wildtype and Te2-knockout mice. **Results:** In COPDGene, individuals with CHIP had a risk of moderate-to-severe and severe or very severe COPD 1.6 and 2.2 times greater than non-carriers, respectively (adjusted 95% confidence intervals [CI], 1.1 to 2.2 and 1.5 to 3.2). These findings were consistently observed in three additional cohorts and meta-analyses of all subjects. CHIP was also associated with decreased FEV_1% predicted in COPDGene (mean between group difference -5.7%; adjusted 95% CI, -8.8 to -2.6), a finding replicated in additional cohorts. Smoke exposure was associated with a small but significant increased risk of having CHIP (OR 1.03 per ten pack-years, 95% CI 1.01-1.05) in the meta-analysis of all subjects. In cigarette smoke exposure models, inactivation of *Tet2* in mouse hematopoietic cells enhanced pulmonary inflammation, increased interferon signaling, decreased TGF-Î² signaling, and increased emphysema development. **Conclusions:** Age-associated somatic mutations in blood cells are associated with the development and severity of COPD, independent of age and cumulative smoke exposure.

PrgmNr 2871 - Exome sequencing and primary care data from the UK Biobank reveal an opportunity for an RNAi therapeutic targeting XDH to treat gout

View session detail

Author Block: **A. Deaton**[1], M. M. Parker[2], L. D. Ward[2], A. Flynn-Carroll[1], R. Hoffing[1], G. Hinkle[1], K. Yucius[1], A. Borodovsky[1], P. Nioi[1]; [1]Alnylam Pharmaceuticals, Cambridge, MA, [2]Alnylam, Cambridge, MA
Disclosure Block: **A. Deaton:** Salary/Employment; Alnylam Pharmaceuticals.
Large biobanks offer the opportunity to identify tractable drug targets through genetic analyses and to examine the real-world efficacy of current therapeutics. We used gene-based collapsing tests to identify genes associated with urate levels in ~364,000 exome-sequenced participants in the UK Biobank. Aggregated predicted LOF (pLOF) and predicted damaging missense variants in XDH associated with lower urate levels (p-value = 3.4 x 10^{-10}, effect = 0.05 SD decrease) and protection from gout (p-value = 0.009, OR = 0.87). XDH encodes xanthine oxidase, an enzyme involved in the production of urate and the target of allopurinol, a drug commonly prescribed to treat gout. We examined allopurinol prescriptions and subsequent urate tests in primary care data from UK Biobank and found that 42% of individuals prescribed allopurinol did not reach target urate levels of 357Âµmol/L. This was highest for those prescribed lower doses of allopurinol, 56% for those with a 100mg prescription compared to 21% for those with a 300mg prescription (p-value= 1.34 x 10^{-80}). This observation is consistent with the known challenges of properly titrating allopurinol in clinical practice to achieve adequate urate control and highlights the need for a novel therapeutic. After confirming high expression of XDH in human liver, we developed ALN-XDH a hepatocyte-directed siRNA targeting XDH. Studies in non-human primates demonstrated durable reduction in XDH protein levels in liver by up to 91% following subcutaneous treatment with ALN-XDH. These results highlight the utility of combining analysis of rare variants with health records to identify drug development opportunities.

PrgmNr 2872 - Gene-environment interactions determine human milk oligosaccharide composition in lactating mothers

View session detail

Author Block: **Z. Y. Fang**[1], S. A. Stickley[1], A. Ambalavanan[1], Y. Zhang[2], B. Robertson[3], C. Yonemitsu[3], K. Miliku[4,5], S. E. Turvey[6,7], P. J. Mandhane[8], E. Simons[5], T. J. Moraes[9], M. R. Sears[10], P. Subbarao[9], L. Bode[3], M. B. Azad[4,5], Q. L. Duan[1,2]; [1]Dept. of BioMed. and Molecular Sci., Queen's Univ., Kingston, ON, Canada, [2]Sch. of Computing, Queen's Univ., Kingston, ON, Canada, [3]Dept. of Pediatrics and Larsson-Rosenquist Fndn. Mother-Milk-Infant Ctr. of Res. Excellence, Univ. of California San Diego, La Jolla, CA, [4]Manitoba Interdisciplinary Lactation Ctr. (MILC), Children's Hosp. Res. Inst. of Manitoba, Winnipeg, MB, Canada, [5]Dept. of Pediatrics and Child Hlth., Univ. of Manitoba, Winnipeg, MB, Canada, [6]Div. of Allergy and Immunology, Dept. of Pediatrics, Univ. of British Columbia, Vancouver, BC, Canada, [7]Dept. of Pediatrics, Child and Family Res. Inst. and British Columbia Childrenâ□□s Hosp., Vancouver, BC, Canada, [8]Dept. of Pediatrics, Univ. of Alberta, Edmonton, AB, Canada, [9]Dept. of Pediatrics, Hosp. for Sick Children and Univ. of Toronto, Toronto, ON, Canada, [10]Dept. of Med., McMaster Univ., Hamilton, ON, Canada

Disclosure Block: **Z.Y. Fang:** None.

Human milk oligosaccharides (HMOs) are carbohydrates secreted exclusively in human breast milk and absent from most infant formulas. HMOs are known to shape the development of the immune system in breastfeeding infants, which have lasting health effects. Earlier studies reported that inter-individual variations in HMO secretion are associated with both genetic and environmental factors. In this study, we determine the impact of interactions between maternal genetics and exposures on HMO composition. Breast milk samples from 980 mothers of the CHILD Cohort Study, collected between 3 to 4 months postpartum, were used to quantify the 19 most abundant HMOs by high-performance liquid chromatography. Genome-wide single nucleotide polymorphisms (SNPs) were genotyped in these mothers using the Illumina HumanCoreExome BeadChip. We applied a generalized linear model to assess the effects of interactions between maternal genetic variants and exposures (e.g., prenatal diet and vitamin intake) or health conditions (e.g., depression, atopy, gestational diabetes) on HMO concentrations. We determined that interactions between SNPs in chromosome 14 and maternal atopy were significantly associated with concentrations of Lacto-N-hexaose (e.g., rs79899738, P = 2.1 Ã□â□□$10^{â□□8}$). We also found that 2â□²-Fucosyllactose is significantly associated with interactions between loci in chromosome 16 and usage of analgesics during labor (e.g., rs12051213, P = 4.3 Ã□â□□$10^{â□□8}$). Finally, interactions between SNPs on chromosome 20 and prenatal fruit consumption significantly modulate 6Â´Sialyllactose concentrations (e.g., rs6036178, P = 2.9 Ã□â□□$10^{â□□9}$). Thus, our study reports interaction effects of maternal genetic and non-genetic factors such as diet, medication use and atopy on HMO composition. A better understanding of the determinants of breast milk composition such as HMOs will help facilitate more nuanced studies of the benefits of breastfeeding on infantsâ□□ health and the development of personalized interventions such as prebiotics.

PrgmNr 2873 - Generalizing Kernel Association Tests with a Protein Structure Based Kernel

View session detail

Author Block: **S. Sadhuka**; Harvard Univ., Cambridge, MA
Disclosure Block: **S. Sadhuka:** None.

Rare coding variants have the potential to identify genes involved in disease. However, few associations have been reported for complex traits, in part due to difficulties in aggregating variants within a gene. One reason is that kernel-based approaches (KATs) â the most widely used rare variant association tests â make strong assumptions about the correlation structure between different variants in a gene, such as independence among variantsâ effects or constant correlation across them. Violations of these assumptions may not only be widespread but predictable from variant properties such as their proximity in 3D protein space.

We hypothesize that (a) functional variants cluster in particular regions of proteins and (b) this clustering is stronger in 3D space than linear sequence space. We verify this hypothesis by showing two types of variant functional effect clustering: (a) variant cellular effects significantly cluster in 3D space (p Based on this hypothesis, we construct a âbiologically informedâ KAT where the correlation between variant effect sizes is inversely proportional to the distance between the variants in 3D space. We show that our test not only mathematically generalizes existing KATs that use non-biologically-informed correlation structures but also â in contrast to prior ad hoc methods â can model how 3D structure influences effects of variants on disease. Our generalization permits many possible variant correlation functions, including those that incorporate biological annotations beyond 3D location.

When used to analyze simulated data in which variant effects are localized to a spherical protein region, we show that our method increased sample sizes by 20-40% relative to existing methods. The increases are strongest when effects are localized to a sphere with radius ~15A, about 20% of an average protein domain. When used to conduct type 2 diabetes (T2D) association analysis within 45,231 exome sequences from the AMP-T2D-GENES consortium, seven of eight known T2D drug targets yielded lower p-values under our methods relative to existing KATs. For instance, for DPP4 our method leverages a cluster of variants to identify a significantly stronger association ($p = 0.006$) than existing KATs ($p = 0.36$).

These results demonstrate that both variant cellular and phenotypic effects exhibit clustering within 3D protein space and provide a foundation to build improved association tests to leverage this information.

PrgmNr 2874 - Genetic associations of protein-coding variants in human disease

View session detail

Author Block: **B. Sun**[1,2], M. Kurki[3,4,5,6], C. N. Foley[7,8], A. Mechakra[9], C-Y. Chen[1], E. Marshall[1], J. Wilk[1], Biogen Biobank Team, M. Chahine[10], P. Chevalier[9], G. ChristÃ©[9], FinnGen, A. Palotie[11,5,4,6], M. J. Daly[11,3,5,4,6], H. Runz[1]; [1]Biogen, Cambridge, MA, [2]BHF Cardiovascular Epidemiology Unit, Dept. of Publ. Hlth.and Primary Care, Univ. of Cambridge, Cambridge, United Kingdom, [3]Psychiatric & Neurodevelopmental Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [4]The Stanley Ctr. for Psychiatric Res., The Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Inst. for Molecular Med. Finland, Univ. of Helsinki, Helsinki, Finland, [6]Analytic and Translational Genetics Unit, Dept. of Med., Massachusetts Gen. Hosp., Boston, MA, [7]Optima Partners, Edinburgh, United Kingdom, [8]Sch. of Clinical Med., Univ. of Cambridge, Cambridge, United Kingdom, [9]UniversitÃ© Claude Bernard Lyon 1, EA4612 Neurocardiology, Lyon, France, [10]UniversitÃ© Laval, Quebec, QC, Canada, [11]Massachusetts Gen. Hosp., Boston, MA

Disclosure Block: **B. Sun:** Salary/Employment; Biogen.

Genome-wide association studies (GWAS) have identified thousands of genetic variants linked to the risk of human disease. However, GWAS have thus far remained largely underpowered to identify associations in the rare and low frequency allelic spectrum and have lacked the resolution to trace causal mechanisms to underlying genes. Here, we combined whole exome sequencing in 392,814 UK Biobank participants with imputed genotypes from 260,405 FinnGen participants (653,219 total individuals) to conduct association meta-analyses for 744 disease endpoints across the protein-coding allelic frequency spectrum, bridging the gap between common and rare variant studies. We identified 975 associations, with more than one-third of our findings not reported previously. Through multiple approaches including computational and functional characterization, we demonstrate population-level relevance for mutations previously ascribed to causing single-gene disorders, map GWAS associations to likely causal genes, explain disease mechanisms, and systematically relate disease associations to levels of 117 biomarkers and clinical-stage drug targets. Our approach benefits considerably from the Finnish genetic background where our theoretical and empirical simulation results suggest the increasing utility of enriched variants for identifying associations quantitatively towards lower allelic frequencies. We identify the most prominent relative power gain in the rarest variant frequency spectrum, highlighting a role for sequencing studies and integrating additional population cohorts with enriched variants for identifying novel disease associations at scale. Combining sequencing and genotyping in two population biobanks allowed us to benefit from increased power to detect and explain disease associations, validate findings through replication and propose medical actionability for rare genetic variants. Our study provides a compendium of protein-coding variant associations for future insights into disease biology and drug discovery.

PrgmNr 2875 - Genetically increased circulating FUT3 level is associated with reduced risk of Idiopathic Pulmonary Fibrosis: a Mendelian Randomization Study

View session detail

Author Block: **T. Nakanishi**[1], A. Cerani[1], V. Forgetta[2], S. Zhou[3], R. J. Allen[4], O. C. Leavy[5], M. Koido[6], D. Assayag[1], R. G. Jenkins[7], L. V. Wain[8], I. V. Yang[9], G. M. Lathrop[1], P. J. Wolters[10], D. A. Schwartz[11], J. B. Richards[12]; [1]McGill Univ., Montreal, QC, Canada, [2]Lady Davis Inst., Montreal, QC, Canada, [3]McGill Univ. Lady Davis Inst., Montreal, QC, Canada, [4]Univ. of Leicester, Leicester, United Kingdom, [5]London Sch. of Hygiene and Tropical Med., London, United Kingdom, [6]Inst. of Med. Sci., The Univ. of Tokyo, Tokyo, Japan, [7]Natl. Inst. for Hlth.Res., Nottingham, United Kingdom, [8]Univ Leicester, Leicester, United Kingdom, [9]Univ Colorado Denver, Aurora, CO, [10]Univ. of California, San Francisico, CA, [11]Natl. Jewish Hlth., Denver, CO, [12]Mc Gill Univ, Montreal, QC, Canada

Disclosure Block: **T. Nakanishi:** None.

Idiopathic pulmonary fibrosis (IPF) is a progressive, fatal fibrotic interstitial lung disease. Few circulating biomarkers have been identified to have causal effects on IPF.

To identify candidate IPF-influencing circulating proteins, we undertook an efficient screen of circulating proteins by applying a two-sample Mendelian randomization (MR) approach with existing publicly available data. For instruments we used genetic determinants of circulating proteins which reside cis to the encoded gene (cis-SNPs), identified by two genome-wide association studies (GWASs) in European individuals (3,301 and 3,200 subjects). We then applied MR methods to test if the levels of these circulating proteins influenced IPF susceptibility in the largest IPF GWAS (2,668 cases and 8,591 controls). We validated the MR results using colocalization analyses to ensure that both the circulating proteins and IPF shared a common genetic signal.

MR analyses of 834 proteins found that a one SD increase in circulating FUT3 and FUT5 was associated with a reduced risk of IPF (OR: 0.81, 95%CI: 0.74-0.88, $p=6.3x10^{-7}$, and OR: 0.76, 95%CI: 0.68-0.86, $p=1.1x10^{-5}$). Sensitivity analyses including multiple-cis SNPs provided similar estimates both for FUT3 (inverse variance weighted [IVW] OR: 0.84, 95%CI: 0.78-0.91, $p=9.8x10^{-6}$, MR-Egger OR: 0.69, 95%CI: 0.50 - 0.97, p=0.03) and FUT5 (IVW OR: 0.84, 95%CI: 0.77-0.92, $p=1.4x10^{-4}$, MR-Egger OR: 0.59, 95%CI: 0.38 - 0.90, p=0.01) FUT3 and FUT5 signals colocalized with IPF signals, with posterior probabilities of a shared genetic signal of 99.9% and 97.7%. Further transcriptomic investigations supported the protective effects of FUT3 for IPF.

An efficient MR scan of 834 circulating proteins provided evidence that genetically increased circulating FUT3 level is associated with reduced risk of IPF.

PrgmNr 2876 - Genetics of male pattern baldness in sub-Saharan Africa

View session detail

Author Block: **J. Lachance**[1], M. Harlemon[1], M. Kim[2], C. N. Simonti[1], R. Janivara[1], T. Rebbeck[3,4], MADCaP Network; [1]Georgia Inst. of Technology, Atlanta, GA, [2]Fred Hutchinson Cancer Res. Ctr., Seattle, [3]Dana-Farber Cancer Inst., Boston, MA, [4]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA
Disclosure Block: **J. Lachance:** None.
Male pattern baldness is a heritable trait, and the prevalence of baldness varies greatly across global populations. Unfortunately, much of what is known about the genetics of male pattern baldness comes from individuals of European descent, and it is unknown if the genetic architecture of this trait varies by ancestry. Here, we analyzed a novel dataset of 1,033 individuals from Ghana, Nigeria, Senegal, and South Africa who were genotyped using the MADCaP Array (a custom genotyping platform that is optimized for African populations). Applying polygenic risk scores for male-pattern baldness to African data, we find that genetic predictions generated from European GWAS perform poorly in sub-Saharan African populations, yielding AUC statistics that are slightly better than random chance. This lack of generalizability is due to multiple causes, including continental differences in allele frequencies, the presence of introgressed Neanderthal alleles in non-African populations, and genotype-by-environment interactions. We subsequently conducted the first African GWAS of androgenic alopecia, finding 24 independent loci that are associated with baldness at age 45 (p-value -5), including one locus that reached genome-wide significance (rs79284602, near *TRIM33* at 1p13.2). Notably, this polymorphism is restricted to African populations. Topologically associated domains near baldness GWAS hits were enriched for genes involved in lipoprotein assembly, remodeling, and clearance pathways. Although the majority of baldness associated loci are autosomal, X-linked variants contribute a relatively large fraction of the SNP-based heritability for this trait in African populations. Collectively, our results highlight the limited transferability of predictions across different ancestries as well as the need to conduct genetic studies in a wide range of populations.

PrgmNr 2877 - Genome-wide association and comparative gene set enrichment analysis identifies both similarities and differences between skeletal growth phenotypes

View session detail

Author Block: **E. Bartell**[1,2], K. Tsuo[3,2,4], W. Gan[5], K. Lin[5], J. B. Cole[6,1,2,7,8], S. Vedantam[7,2], R. M. Salem[9], R. G. Walters[5], J. N. Hirschhorn[7,1,2]; [1]Dept. of Genetics, Harvard Med. Sch., Boston, MA, [2]Program in Med. and Population Genetics, Broad Inst. of Harvard and MIT, Cambridge, MA, [3]Harvard Univ., Cambridge, MA, [4]Massachusetts Gen. Hosp. Res. Inst., Boston, MA, [5]Wellcome Trust Ctr. for Human Genetics, Oxford, United Kingdom, [6]The Broad Inst. of MIT and Harvard, Cambridge, MA, [7]Div. of Endocrinology and Ctr. for Basic and Translational Obesity Res., Boston Children's Hosp., Boston, MA, [8]Program in Metabolism, Broad Inst. of MIT and Harvard, Cambridge, MA, [9]Herbert Wertheim Sch. of Publ. Hlth., UC San Diego, San Diego, CA

Disclosure Block: **E. Bartell:** None.

Background: Genome-wide association studies (GWAS) have identified thousands of common genetic variants associated with human height; however, less is known about the genetics of body proportion. Understanding the genetics of Sitting Height Ratio (SHR), leg length (Leg), and sitting height (SH), commonly-used measures of skeletal proportion, would further the study of skeletal growth.
Objective: To understand skeletal growth using genetic studies of SHR and height-related phenotypes.
Design/Methods: We use data from the UK Biobank (UKB) and China Kadoorie Biobank (CKB) to expand our genome-wide association study of SHR (Chan et al. 2015) from ~25,000 to >500,000 individuals from 2 ancestries, and analyze Leg and SH. We perform GWAS using BOLT-LMM in ~400,000 european and ~76,000 east asian individuals. MAGMA (de Leeuw 2015) was used to perform Gene Set Enrichment Analysis (GSEA) on individual traits and on heterogeneity summary statistics calculated using Cochranâ□□s Q (Cochran 1937).
Results: We identified between 900-1800 genome-wide significant SHR-, Height-, Leg-, and SH-associated lead variants (pConclusions: Using GWAS in two major ancestries, we identified >3100 novel associations with SHR, Leg, and SH. We performed conditional analysis and developed and implemented comparative-GSEA to provide insights into the genetic and biological differences between measures of body proportion and overall skeletal growth.

PrgmNr 2878 - Genome-wide inter-chromosomal epistatic associations identified across complex diseases in the ~300,000 participants from eMERGE and UK Biobank

View session detail

Author Block: **S. S. Verma**[1], P. Singhal[2,3], A. Lucas[4,3], Y. Veturi[5,3], C. Weng[6], s. pendergrass[7], I. J. Kullo[8], S. J. Schrodi[9,10], D. Fasel[11], D. J. Schaid[8], O. Dikilitas[8,12], P. M. Sleiman[13], H. Hakonarson[14], M. D. Ritchie[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Philadelphia, MA, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]411 Waupelani Drive A-341, Philadelphia, PA, [5]State College, PA, [6]Columbia Univ., New York, NY, [78]Mayo Clinic, Rochester, MN, [9]Marshfield Clinical Res Fndn, Marshfield, WI, [10]Univ. of Wisconsin, Madison, WA, [11]New York, NY, [12]Children's Hosp. of Philadelphia, Philadelphia, PA, [13]CHOP, Philadelphia, PA, [14]Children S Hosp. of Philadelphia, Philadelphia, PA

Disclosure Block: **S.S. Verma:** None.

Capitalizing on linkage disequilibrium (LD) patterns to determine population substructure allows the discovery of additive association signals in genome-wide association studies (GWAS). As standard GWAS analyses are well-powered to interrogate additive models, investigating other plausible modes of inheritance such as dominance and epistasis may often require new approaches. Epistasis - interaction between genes, play an essential role in elucidating complex genetic networks that impact the human genome's structure and evolution. Given that evidence from model organism studies indicates long-range high LD regions to be under evolutionary selection, we hypothesized that these regions might play a key role in regulating disease mechanisms across various complex traits. Thus, we selected 20 diverse complex traits (neurological, ocular, cardiometabolic, immune) and created case/control cohorts of individuals in the eMERGE and UKBB datasets. We calculated Ohtaâ□□s D statistic on genome-wide pairs of variants to identify pairs in long-range (> 0.25cM) LD due to epistatic selection. On these resulting 136,019 SNPs culminating in a total of 5,625,845 SNP-SNP models, we used a penalized regression framework to determine the association of SNP-SNP pairs with a disease. Across 12 of the 20 phenotypes, we found 290 models (majority inter-chromosomal) that replicated in UKBB and eMERGE after multiple hypothesis testing correction. Characterization of replicating models indicate the genes they map to are 1) highly conserved gene families with complex roles in multiple pathways, 2) essential genes, and/or 3) already associated in the literature with complex traits that have diverse phenotypic presentations. Together, these results demonstrate the highly pleiotropic nature of variants under epistatic selection in conserved regions, supporting the hypothesis that epistatic interactions regulate diverse clinical mechanisms and can produce a range of phenotypic outcomes. A key example is the WFS1 gene, coding for the calcium-regulating wolframin protein. Over 200 mutations in WFS1 have been associated with Wolframin Syndrome, characterized by different combinations of diabetes, optic atrophy, urinary tract dysfunction, loss of hearing, and neurological symptoms. Our results support the conclusion that epistatic interactions between *WFS1* and other genes, including *VRK2, NT5C2, INA, and DIP2C*, regulate this syndrome and produce the range of phenotypes seen. This study examines numerous such cases and sheds light on the pleiotropic interactions underlying disease etiologies providing a more unified view of epistasis and pleiotropy in complex traits.

PrgmNr 2880 - Human milk oligosaccharides alter the milk microbiota in lactating mothers and impact the gut microbiota of breastfed infants with asthma

View session detail

Author Block: **S. A. Stickley**[1], Z. Y. Fang[1], A. Ambalavanan[1], Y. Zhang[2], J. Choi[1], B. Robertson[3], C. Yonemitsu[3], K. Miliku[4,5], S. E. Turvey[6,7], P. J. Mandhane[8], E. Simons[5], T. J. Moraes[9], M. R. Sears[10], P. Subbarao[9], L. Bode[3], M. B. Azad[4,5], Q. L. Duan[1]; [1]Dept. of BioMed. and Molecular Sci., Queen's Univ., Kingston, ON, Canada, [2]Sch. of Computing, Queen's Univ., Kingston, ON, Canada, [3]Dept. of Pediatrics and Larsson-Rosenquist Fndn. Mother-Milk-Infant Ctr. of Res. Excellence, Univ. of California San Diego, La Jolla, CA, [4]Manitoba Interdisciplinary Lactation Ctr. (MILC), Children□□s Hosp. Res. Inst. of Manitoba, Winnipeg, MB, Canada, [5]Dept. of Pediatrics and Child Hlth., Univ. of Manitoba, Winnipeg, MB, Canada, [6]Div. of Allergy and Immunology, Dept. of Pediatrics, Univ. of British Columbia, Vancouver, BC, Canada, [7]Dept. of Pediatrics, Child and Family Res. Inst. and British Columbia Children□□s Hosp., Vancouver, BC, Canada, [8]Dept. of Pediatrics, Univ. of Alberta, Edmonton, AB, Canada, [9]Dept. of Pediatrics, Hosp. for Sick Children and Univ. of Toronto, Toronto, ON, Canada, [10]Dept. of Med., McMaster Univ., Hamilton, ON, Canada
Disclosure Block: **S.A. Stickley:** None.

Human milk consists of hundreds of complex glycans known as human milk oligosaccharides (HMOs), which are absent from most infant formulas and associated with changes in the milk microbiota of lactating mothers and the gut microbiota of breastfed infants. In addition, HMOs contribute to the development of the immune system in infants and have been associated with infant health outcomes. The aim of our study is to generate new hypotheses for the underlying mechanisms by which HMOs impact the milk and gut microbiota to influence lung health in breastfed infants. We quantified the 19 most abundant HMOs using high performance liquid chromatography from the human milk samples of 980 mothers of the CHILD cohort study, collected at 3 months postpartum. A subset of these human milk samples (N=721) were used to assess milk microbial composition using 16S rRNA sequencing. Gut microbial composition were determined by 16S rRNA sequencing of stool samples from breastfed infants at 3 months (N=323) and at 1 year (N=342). We performed differential abundance analyses to determine how HMO concentrations modify the mothers□□ milk microbiota and the infants□□ gut microbiota. Our results indicate that high concentrations of the HMO lacto-N-neotetraose is associated with elevated levels of *Lactobacillus iners* in the human milk (P = 7.23 x 10-47). In turn, breastfed children exposed to high *L. iners* in human milk had lower prevalence of recurrent wheeze (P = 1.11 x 10-26) and asthma (P = 1.25 x 10-6). Moreover, specific HMOs such as difucosyllactose (P = 1.20 x 10-30) and 3□²-sialyllactose (P = 1.96 x 10-30) were associated with lower abundance of *Bacteroides caccae* in the gut at 3 months. Children with lower *B. caccae* in the gut microbiota had lower prevalence of recurrent wheeze (P = 4.40 x 10-14) and asthma (P = 2.07 x 10-11). In conclusion, our study suggests that HMO composition among lactating mothers influences the milk and gut microbiotas, which in turn, impacts the risk of recurrent wheeze and asthma among breastfed infants.

PrgmNr 2881 - Identifying synthetic interactions between synonymous mutations in the UK Biobank WES data using REVEAL: Biobank

View session detail

Author Block: **Z. Pitluk**; Pardigm4, Inc., Waltham, MA
Disclosure Block: **Z. Pitluk:** None.

Historically, most research into discovering the impact of mutations on human health and disease has focused on algorithmically determined mutations of consequence. However, synonymous or silent mutations are known to have a dramatic impact on the rate of translation of proteins, and potentially on the stability of individual proteins and complexes of proteins. Multiple studies have linked synonymous mutations to changes in protein expression, and translation efficiency, among others. In this work, we evaluate the role of synonymous mutations across the 200K whole exome sequencing data made available through the UK Biobank (Application ID: 51518) by systematically looking for evidence of synthetic interactions between individual synonymous mutations in pairs of genes. To perform this large-scale study, we will be using Paradigm4â□□s software analytics platform, REVEAL™: Biobank, which is built upon an array-native computational engine called SciDB. We will employ fisherâ□□s exact test and check for co-occurrence and anti-occurrence of mutations to efficiently rank putative synthetic lethal and synthetic obligatory interactions. Additionally, we will utilize linkage disequilibrium and burden tests to measure correlation between variant pairs and their effect on phenotypes. Finally, we will use bioinformatics to develop testable hypotheses to understand prospective interactions.

PrgmNr 2882 - Incident disease associations with mosaic chromosomal alterations on autosomes, X and Y chromosomes: insights from a phenome-wide association study in the UK Biobank

View session detail

Author Block: **S-H. Lin**[1], D. W. Brown[1], B. E. Rose[2], F. R. Day[3], O. Lee[1], S. Khan[1], J. Hislop[1], S. J. Chanock[4], J. R. Perry[5], M. J. Machiela[1]; [1]Natl. Cancer Inst., Rockville, MD, [2]Univ. of Miami, Miami, FL, [3]Cambridge Univ, Cambridge, United Kingdom, [4]Natl. Cancer Inst, Rockville, MD, [5]Univ Cambridge, Cambridge, United Kingdom

Disclosure Block: **S. Lin:** None.

Mosaic chromosomal alterations (mCAs) can develop in hematopoietic cells and achieve a high clonal fraction of circulating cells: events can be classified as large chromosomal gains, losses and copy-neutral losses of heterozygosity (LOH). The clonal selection of mutated cells harboring mCAs relative to normal cells can alter gene dosage and expression, potentially having an impact on leukocyte function. While many individuals with mCAs have no apparent adverse impacts on health at the time of assay, the effect of such events can be explored in prospective studies, such as the UK Biobank. We performed a phenome-wide association study (PheWAS) using existing genotype and phenotype data from 482,396 UK Biobank participants to investigate potential associations between mCAs and incident disease. A total of 17,113 (3.5%) participants had at least one detectable autosomal event, 43,297 (19.6%) males had mosaic Y loss and 12,550 (4.8%) females had mosaic X loss. First occurrence of 1,764 ICD-10-coded diseases were derived from linkage to primary care, hospital admission, cancer registry, death register data as well as self-report disease history. PheWAS models included adjustment for age, age^2, sex, and a detailed 25-level smoking variable. Our adjusted analysis identified a total of 50 incident disease outcomes associated with mCAs at PheWAS significance levels (P-5). Overall autosomal mCA associations included lymphoid and myeloid malignancies, infectious disease risk and cancer related phenotypes. When a PheWAS was performed separately for each autosome, we observed outcomes associated with mCAs for nearly every chromosome (e.g., chronic lymphoid leukemia) whereas other associations were restricted to fewer chromosomes (e.g., myeloid leukemia). For sex chromosome analyses, we observed mCAs in chromosome X were associated with increased lymphoid leukemia risk and mCAs of chromosome Y were linked to potential reduced metabolic disease risk. Further exploratory analyses using prevalent disease as well as reported medication data further confirm reported mCA associations. Our findings detail a broad spectrum of associations between mCAs and incident diseases that varies by type of mCA and highlight the critical importance of careful covariate adjustment to minimize confounding.

PrgmNr 2883 - Integration of a *MUC5B* Promoter Variant and a Polygenic Risk Score for Idiopathic Pulmonary Fibrosis and Interstitial Lung Abnormalities

View session detail

Author Block: **M. Moll**[1], S. Chun[2], R. J. Allen[3], A. J. Ghosh[1], J. Ziniti[1], R. K. Putman[1], H. Hatabu[1], J. Hecker[4], B. D. Hobbs[5], E. K. Silverman[6], B. A. Raby[7], L. V. Wain[8], G. M. Hunninghake[6], M. H. Cho[9]; [1]Brigham and Women's Hosp., Boston, MA, [2]Boston Children's Hosp., Boston, MA, [3]Univ. of Leicester, Leicester, United Kingdom, [4]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, Germany, [5]Needham, MA, [6]Brigham & Women's Hosp, Boston, MA, [7]Newton Center, MA, [8]Univ Leicester, Leicester, United Kingdom, [9]Duxbury, MA

Disclosure Block: **M. Moll:** None.

Rationale: Idiopathic pulmonary fibrosis (IPF) is characterized by progressive lung scarring and death. Early disease detection is paramount. Chest computed tomographic imaging can detect interstitial lung abnormalities (ILA), which can progress to IPF. The *MUC5B* promoter variant rs35705950 is a common variant (MAF [Europeans]=0.11) of large effect size that confers risk for both IPF (OR 6-13, heterozygotes) and ILA (OR 2-3). The relative contributions of rs35705950 and other variants to IPF and ILA are not clear. We hypothesized that a polygenic risk score (PRS) excluding the *MUC5B* region would add complementary predictive value to the *MUC5B* variant for IPF and ILA. Methods: Using previously published European-ancestry genome-wide association summary statistics for IPF, we used lassosum to develop an IPF PRS excluding a 500 kb region around rs35705950 (PRS-NO-M5B). We trained a composite risk score of the PRS-NO-M5B and *MUC5B* rs35705950 (PRS+M5B) using 10-fold cross-validation (100 bootstraps) in IPF cases and controls from the external Lung Tissue Research Consortium (LTRC). We tested the associations of PRS-NO-M5B, rs35705950 alone, and PRS+M5B with IPF in LTRC and ILA in the Genetic Epidemiology of COPD (COPDGene) study. Multivariable logistic regressions were performed, adjusting for age, sex, smoking, and genetic ancestry. We assessed interaction by testing association of cross-product terms with each outcome. Area-under-the-curve (AUC) analyses were used to assess model predictive performances. We examined phenotypic variance explained by each risk factor using Nagelkerke R^2. Results: In LTRC (270 controls, 255 IPF cases), the PRS-NO-M5B (OR 4.5 [95% CI: 3.3 - 6.1], p=8.0e-22) and rs35705950 (OR 3.6 [95% CI: 2.4 - 5.2], p=6.4e-11) were both associated with IPF, and when considered in a single model, effect sizes were largely unchanged. PRS+M5B demonstrated a larger effect (OR 18 [95% CI: 10 - 30], p=5.0e-26) compared to either component risk factor. The PRS-NO-M5B X rs35705950 interaction term was not significant (p=0.7). There was a stepwise increase in predictive performances of models including rs35705950 (AUC 0.74), PRS-NO-M5B (AUC 0.83), and PRS+M5B (AUC 0.86). Phenotypic variability in IPF explained by PRS+M5B (R^2=0.41) was greater than PRS-NO-M5B (R^2=0.33) and rs35705950 alone (R^2=0.12). In COPDGene (6417 controls, 250 ILA), PRS+M5B (OR 1.5 [95% CI: 1.2-1.9],p=5.4e-4) and rs35705950 (OR 1.8 [95% CI: 1.4-2.4], p=2.7e-6), but not the PRS-NO-M5B (p=0.08), were significantly associated with ILA. Conclusions: The *MUC5B* rs35705950 variant and a PRS excluding the *MUC5B* region offered complementary predictive value for IPF, but not ILA.

PrgmNr 2884 - Investigating the transferability of European height associated loci in Africans

View session detail

Author Block: **D. Ju**[1], M. Hansen[2], D. N. Harris[3], J. B. Hirbo[4], E. Mbunwe[2], A. Ranciaro[1], S. R. Thompson[2], D. Woldemeskel[5], C. Zhang[1], W. Beggs[1], G. Belay[5], C. Fokunang[6], G. Mokone[7], S. Mpoloka[7], A. K. Njamnshi[6], T. Nyambo[8], Trans-Omics for Precision Medicine (TOPMed) Consortium, S. Omar[9], I. Mathieson[10], S. A. Tishkoff[11]; [1]Perelman Sch. of Med., Philadelphia, PA, [2]Univ Pennsylvania, Philadelphia, PA, [3]Univ. of Pennsylvania, Wynnewood, PA, [4]Vanderbilt Univ. Sch. of Med., Nashville, TN, [5]Addis Ababa Univ., Addis Ababa, Ethiopia, [6]Univ. of Yaounde I, Yaounde, Cameroon, [7]Univ. of Botswana, Gaborone, Botswana, [8]Kampala Intl. Univ., Kampala, Uganda, [9]Kenya Med. Res. Inst., Nairobi, Kenya, [10]Univ. of Pennsylvania, Philadelphia, PA, [11]Univ. of Pennsylvania Sch. of Med., Philadelphia, PA

Disclosure Block: **D. Ju:** None.

African populations have been underrepresented in genome-wide association studies (GWAS), limiting our understanding of the genetic architecture of complex traits. We studied standing height in a Sub-Saharan African cohort (N=2274) that was whole-genome sequenced as part of the TOPMed initiative. This cohort represents diverse ancestry groups in Sub-Saharan Africa, including Nilotic pastoralist and Central African hunter-gatherer populations, that exhibit phenotypic extremes in height. We performed GWAS and meta-analysis with GWAS of other African ancestry cohorts to examine the transferability of height loci across populations and evaluated polygenic score prediction for diverse African ancestries. For the GWAS we used a linear mixed model, controlling for population stratification with a genetic relatedness matrix. By permuting phenotypes within ethnic groups, we found this approach sufficiently controlled for inflation for common variants above 5% minor allele frequency in our heterogenous Sub-Saharan African dataset. To increase power, we meta-analyzed our GWAS with previously published GWAS of African cohorts (N=43,984). We observed at least 15 independent and significant loci, some of which replicate known height associated loci such as *HMGA2* and *LCORL.* To examine the transferability of European height associations in Africans, we looked at concordance in direction of effect from significant GWAS variants from a European cohort from the UK biobank (2097 SNPs) and observed 64.5% concordance (P = 1.59Î¾10-40). Stratifying by the top tenth percentile of most associated SNPs in the African ancestry meta-analysis, we observed 90.5% concordance (P = 3.08Î¾10-35). Given this general consistency in effects, we investigated how well these European height associated variants predict height in the Sub-Saharan African cohort using polygenic scores (PS). We observed a partial R^2 of 0.048 for the PS when controlling for age and sex, higher than what has been previously reported in other African cohorts. In a model controlling for genetic ancestry, the partial R^2 of the PS was 0.033, indicating the PS explained some of the variance in height across ancestry groups. When comparing variation in mean height across ethnic groups with mean polygenic scores, however, we found that European height alleles did not explain these population differences for those at the extremes of the distribution in Africa.

PrgmNr 2885 - Large-scale exome sequencing identifies 8 genes impacting adult cognitive function

View session detail

Author Block: **C-Y. Chen**[1], R. Tian[1], J. Z. Liu[2], M. Lam[3], Biogen Biobank Team, E. A. Tsai[1], H. Huang[3], T. Lencz[4], T. Ge[5], H. Runz[1]; [1]Biogen, Cambridge, MA, [2]GSK, Philadelphia, PA, [3]The Broad Inst., Cambridge, MA, [4]Zucker Hillside Hosp., Glen Oaks, NY, [5]Massachusetts Gen. Hosp., Boston, MA
Disclosure Block: **C. Chen:** Salary/Employment; Biogen.
Genome-wide association studies have identified over one thousand genetic loci of relevance to cognitive function in adults. However, the contribution of rare coding variants to cognitive function have largely remained unknown. Here, we present the first large-scale exome sequencing study investigating the impact of protein-coding variants on cognitive phenotypes in the adult general population. Leveraging whole-exome sequencing data from 454,787 UK Biobank (UKB) participants ascertained for educational attainment (EDU), reaction time (RT) and verbal-numerical reasoning (VNR), we discovered that exome-wide protein truncating variant (PTV) and missense burden have significant damaging effects on cognitive function. This was reflected in lower educational attainment, longer reaction time and lower verbal-numerical reasoning score in carriers of rare coding variants (exome-wide PTV burden: P=1.95x10^{-21} for EDU, 8.79x10^{-19} for RT and 6.99x10^{-22} for VNR; missense burden: P=5.95x10^{-24} for EDU, 5.95x10^{-4} for RT and 4.87x10^{-12} for VNR). Furthermore, the strongest signals were driven by PTVs and damaging missense variants (MPC>3 and 3â¥MPC>2) in loss of function intolerant genes (pLIâ¥0.9), indicating potential impact from both classes of variants on cognitive function. Through gene-based PTV burden tests, we identified 8 adult cognitive function genes (*ADGRB2*, *KDM5B*, *GIGYF1*, *ANKRD12*, *SLC8A1*, *RC3H2*, *CACNA1A* and *BCAS3*) at Bonferroni corrected exome-wide significance level, and 5 additional genes at FDR *Q*NDUFA6, *ARHGEF7*, *C11orf94*, *KIF26A*, and *MAP1A*). For established Mendelian disease genes, we showed that the heterozygous PTVs are associated with a milder form of cognitive impairment in adults, highlighting the spectrum of cognitive phenotypes and related disorders that could be influenced by PTVs. We further contrasted cognitive function genes with genes underlying developmental delay and autism spectrum disorder and discovered shared genetic mechanisms that influence cognitive function at different life stages. Gene sets identified through PTV-based burden analyses strengthened prior evidence that distinct biological processes including neurogenesis and synapse formation modulate cognitive function. Finally, we demonstrated additive contributions of rare coding variants and common variant-based polygenic risk on cognitive function in adults. Our findings uncovered a substantial contribution of rare protein-coding variants on cognitive function, identified novel cognitive genes, and provided new insights into the genetics of cognitive function in relation to neurodevelopmental disorders and developmental delay.

PrgmNr 2886 - Neuroimaging-guided functional variant multi-ancestry PheWAS using electronic health records from UK Biobank

View session detail

Author Block: **Y. Veturi**[1], C. Davatzikos[1], M. D. Ritchie[2]; [1]Univ. of Pennsylvania, State College, PA, [2]Univ. of Pennsylvania, Philadelphia, PA
Disclosure Block: **Y. Veturi:** None.
Several studies have identified genetic overlaps between neuroimaging data and complex human diseases (e.g. Alzheimerâ□□s disease, type II diabetes, stroke, etc.). A comprehensive investigation of functional associations connecting neuroimaging phenotypes to diseases in electronic health records (EHR) **across the phenome** can yield novel image-derived phenotype (IDP) biomarkers for complex diseases as well as provide pleiotropic gene targets for drug repurposing. We first conducted a phenome-wide imaging study (PheWIS) in UK Biobank to identify disease-IDP associations (*n*=40,201 samples) between 689 International Classification of Disease (ICD) codes and 870 structural and diffusion MRI IDPs. Next, we conducted a phenome-wide association study (PheWAS) across 689 ICD codes (*n*=452,595 samples) on 52,570 Bonferroni-significant genetic variants (p-valuers34404554/*TOMM40*: Alzheimerâ□□s disease and angina pectoris), we also found pleiotropic associations between IDPs and diseases across the phenome including respiratory (e.g. *ORMDL3*: asthma), digestive (e.g. *GBAP1*: diseases of stomach and duodenum), musculoskeletal (e.g. *PSMB9*: rheumatoid arthritis), eye and adnexa (e.g. *CDKN2B*: glaucoma), genitourinary (e.g. *SP1*: hydrocele and spermatocele) systems as well as neoplasms (e.g. *HLA-DQB1*: follicular non-Hodgkinâ□□s lymphoma). We will conduct statistical colocalization as well as obtain causal pathways between correlated IDP clusters and complex diseases in independent medical biobanks to robustly identify endophenotypes for future prevention and treatment strategies.

PrgmNr 2887 - Organelles and aging: a human genetics approach

View session detail

Author Block: **R. Gupta**[1,2], K. J. Karczewski[1], D. Howrigan[1], B. M. Neale[3,1], V. K. Mootha[3,1]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Harvard Med. Sch., Boston, MA, [3]Massachusetts Gen. Hosp., Boston, MA

Disclosure Block: **R. Gupta:** None.

Most age-related human diseases are accompanied by a decline in cellular organelle integrity, including impaired lysosomal proteostasis, elevated endoplasmic reticular stress, and defective mitochondrial oxidative phosphorylation. While classical inborn errors of metabolism tend to be mapped to â□□causalâ□□ organelles (e.g., lysosomal storage diseases), an open question is the degree to which inherited variation impacting each organelle contributes to common age-related disease pathogenesis.

Here, we evaluate if organelle-relevant loci confer greater-than-expected age-related disease risk. As mitochondrial dysfunction is a â□□hallmarkâ□□ of aging, we begin by comprehensively assessing loci relevant to mitochondria: nuclear DNA loci near genes producing mitochondria-localizing proteins, published nuclear DNA quantitative trait loci for biomarkers of mitochondrial dysfunction, and common mitochondrial DNA variants. To our surprise, we observe a lack of enrichment across 24 age-related traits using several distinct approaches. Within nine other organelles, we find no enrichment with one exception: the nucleus. Via further partitioning of the ~6300 genes annotated to contribute to the nuclear proteome, we find that this signal primarily emanates from nuclear transcription factors. By sub-dividing transcription factors across dimensions of DNA-binding domain, age, and breadth of gene expression across human tissues, we found signal predominantly among the non-KRAB domain-containing transcription factors. Further analyses of genetic loci associated with parental lifespan and healthspan continued to prioritize nuclear transcription factors over other organelles. In agreement with our results, we find that genes encoding several organelles tend to be â□□haplosufficient,â□□ while we observe strong purifying selection against protein-truncating variants impacting the nucleus.

Our work identifies common variation near nuclear transcription factors as having outsize influence on age-related trait risk, motivating future efforts to determine if and how this variation contributes to age-related organelle deterioration.

PrgmNr 2888 - Pakistan Genomic Resource: The worldâ□□s largest biobank of human knockouts

View session detail

Author Block: **D. Saleheen**[1], S. Khalid[2], A. Rasheed[3], P. Natarajan[4], S. Kathiresan[4], A. V. Khera[4], P. Frossard[3], Pakistan Genomic Resource; [1]Columbia Univ., New York, New York, NY, [2]Columbia Univ. Irving Med. Ctr., NY, NY, [3]Ctr. for Non-Communicable Diseases, Karachi, Pakistan, [4]Massachusetts Gen. Hosp., Boston, MA

Disclosure Block: **D. Saleheen:** Grant/Contracted Research Support (External); Regeneron Pharmaceuticals, Eli-Lilly, GSK, Novartis, Novo-Nordisk, NGM, Variant Bio, Genentech, Astra Zeneca.

Individuals carrying 2 copies of plof (putative loss of function) variants, disrupting both copies of a gene (â□□human knock-outsâ□□) can provide valuable insights into gene function and help assess the therapeutic potential of pharmacological inhibition of a gene target. The chances of seeing homozygous plof carriers in consanguineous individuals, whose parents are cousins, are much higher compared to the general population. The Pakistan Genomic Resource (PGR) is an ongoing study that will sequence 1 million participants living in Pakistan, a region of the world with high levels of consanguinity. We have currently enrolled >150,000 participants, 40% of which report that their parents are second cousins or closer. After applying standard plof filtering on the initial 77,929 whole-exome / whole-genome sequenced samples, we identify 15,840 (20.3% of total population) homozygous carriers of rare plofs (alternate allele frequency SLC30A8 was associated with reduced risk of type 2 diabetes, at *GDF15* with absent growth differentiation factor 15 (GDF15) levels and homozygosity with MHCR1 was associated with an increase in obesity risk. Given projections based on our data and others, sequencing up to 1 million participants would allow us to observe up to 90% of all viable knock outs for genes in which no homozygous plof has been observed yet, resulting in a unique resource to help understand novel biology and inform drug discovery efforts.

PrgmNr 2890 - Polygenic gene-environment interactions: from genetic architecture to pharmacogenomics

View session detail

Author Block: **A. R. Marderstein**[1], L. Poyraz[2], S. Kulm[1], C. Peng[3], R. Tamimi[1], O. Elemento[1], A. G. Clark[2]; [1]Weill Cornell Med., New York, NY, [2]Cornell Univ., Ithaca, NY, [3]Brigham and Womenâ□□s Hosp., Boston, MA

Disclosure Block: **A.R. Marderstein:** None.

While genetic effects are often strongly modulated by the environment in model organisms, single-locus gene-environment (GxE) interactions have been a challenge to detect in humans. On the other hand, polygenic GxE interactions, which involve interactions with thousands of small-effect causal variants across the genome, may play an underappreciated role in the genetic architecture of many human phenotypes. To explore this, we analyzed body mass index levels (BMI) and breast cancer (BC) in UK Biobank (UKB).

For BMI: We find that the polygenic contribution is strongly modulated by lifestyle factors. While only the *FTO* locus contained single-SNP interactions reaching genome-wide significance (*P* -8), the three tested GxE interactions between a BMI polygenic score (PGS) and physical activity (PA), sedentary behavior, and alcohol intake are each highly significant (*P* -15). For PA, the effect size of the BMI PGS in low-PA individuals is nearly 34% greater than individuals with high-PA. In a stratified GWAS of only high-PA individuals, we found that individual SNP effects are reduced by 37% on average compared to SNP effects estimated in only low-PA individuals; generally, this aligns with the estimated difference in SNP-heritability between low- and high-PA individuals (35%) despite nearly identical genetic correlations (r_g=0.95). Importantly, individuals with a PGS in the highest risk decile (>90th percentile) and high-PA levels have the same mean BMI as those with an average PGS (40-60th percentile) but low-PA levels. Thus, we found widespread evidence of GxE polygenicity with important implications regarding BMI genetic architecture and prediction.

For BC: We used a PGS-based approach to explore whether widely-taken approved drugs (not targeting BC) modulate BC genetics in UKB women. Notably, we find that a BC PGS explains nearly three-times greater variation in disease risk within corticosteroid users compared to non-users. We show that the PGS can be used to discover the underlying SNP interactions, mapping 35 genes significantly interacting with corticosteroid use (*FDR* HR = 3.41 per-allele within users). In comparison, there are no differences in BC risk within the reference allele homozygotes.

Overall: The high prevalence of polygenic GxE has great potential, from improving risk prediction to analyzing off-target drug effects.

PrgmNr 2891 - Prioritizing Research Variants in the NIH Undiagnosed Diseases Program

View session detail

Author Block: **D. R. Adams**[1], B. N. Pusey[2], C. J. Tifft[3], C. Toro[4], M. Malicdan[4], T. C. Markello[5], W. A. Gahl[6], Undiagnosed Diseases Network; [1]NIH, Bethesda, MD, [2]Natl Human Genome Res. Inst., Bethesda, MD, [3]Office of the Clinical Director, Bethesda, MD, [4]NIH, Bethesda, MD, [5]NIH - NHGRI, Bethesda, MD, [6]NHGRI (NIH), Bethesda, MD

Disclosure Block: **D.R. Adams:** None.

A significant number of rare, high-penetrance, heritable diseases remain undiagnosed after standard of care clinical exome and genome analysis. Follow up research analysis occasionally yields unusual or hard to detect mutations affecting known disease genes. For remaining cases, undiagnosed disease research requires identification and prioritization of DNA variants in potential new disease genesâ□□'research variantsâ□□. The selection of research variants must balance an estimate of prior probability of success, research resources, bioinformatic assessments, gene function correlation with patient phenotypes and availability of expertise. Optimally, selection should be designed to minimize ascertainment bias based on reviewer expertise, quantity of available gene information and low-evidence prior assessments of disease-gene associations. The NIH Undiagnosed Diseases Program, a clinical site of the NIH Undiagnosed Diseases Network, was founded in 2008 to investigate and diagnose diseases that remain undiagnosed after extensive medical workup. Of 1500 evaluated individuals (including family members), there were 1032 *probands* who remained without a conclusive diagnosis at the time of our in-person clinical evaluation. For these individuals, diagnoses have been made or are under validation for approximately 281 (27%). Of the 751 undiagnosed individuals, 680 remain under active investigation. Of the 680 ongoing cases, 303 have been subject to exome or genome sequencing, with 185 (61%) having one or more identified research variants under study. Past UDP research variants have been converted to diagnoses via a variety of mechanisms including published assertion of a new disease by the UDP or UDN, recognition after external publication of a new disease assertion, application of new bioinformatics analyses, published phenotype expansions and other events. Conversely, research variants have been deprioritized based on laboratory research data, public database updates and other sources of new information. We present data summarizing and describing the patterns observed for past UDP research variants and describe our experience-based considerations for an optimized prioritization strategy.

PrgmNr 2892 - Projecting genetic associations and drug transcriptional profiles through gene expression patterns reveal disease etiology and potential mechanisms for therapeutic strategies

View session detail

Author Block: **M. Pividori**[1], S. Lu[2], B. Li[3], C. Su[2], M. E. Johnson[2], W-Q. Wei[4], Q. Feng[4], B. Namjou[5], K. Kiryluk[6], I. J. Kullo[7], Y. Luo[8], B. D. Sullivan[9], C. Skarke[10], M. D. Ritchie[1], S. F. A. Grant[1,2,11], C. S. Greene[12]; [1]Dept. of Genetics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [2]Ctr. for Spatial and Functional Genomics, Div. of Human Genetics, Childrenâ□□s Hosp. of Philadelphia, Philadelphia, PA, [3]Dept. of BioMed. Data Sci., Stanford Univ., Stanford, CA, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]Children's Hosp. Med. Ctr., Cincinnati, OH, [6]Columbia Univ., New York, NY, [7]Mayo Clinic, Rochester, MN, [8]Northwestern Univ., Chicago, IL, [9]Sch. of Computing, Univ. of Utah, Salt Lake City, UT, [10]Inst. for Translational Med. and Therapeutics, Dept. of Med., Univ. of Pennsylvania, Philadelphia, PA, [11]Dept. of Pediatrics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [12]Ctr. for Hlth.AI, Univ. of Colorado Sch. of Med., Aurora, CO

Disclosure Block: **M. Pividori:** None.

Human diseases have tissue-specific etiologies and manifestations. In this context, determining how genes influence these complex phenotypes requires mechanistically understanding expression regulation across different cell types, which in turn should lead to improved treatments. Integrating functional and GWAS data has improved the identification of these transcriptional mechanisms which, when dysregulated, commonly result in tissue- and cell lineage-specific pathology. However, widespread gene pleiotropy and polygenic traits reveal the highly interconnected nature of transcriptional networks, which complicates the interpretation of genetic effects and hampers translational efforts. We have developed a polygenic approach that maps both gene-trait associations and drug-transcriptional responses into a common representation based on tissue-specific gene co-expression patterns.

We integrated thousands of gene-trait associations (using TWAS from PhenomeXcan) and transcriptional profiles of drugs (LINCS L1000) into a low-dimensional representation learned from public gene expression data on tens of thousands of RNA-seq samples (recount2). This low-dimensional space comprised features representing groups of genes (gene modules) with coordinated expression across different tissues and cell types. When mapping gene-trait associations to this reduced expression space, we observed that diseases were significantly associated with gene modules expressed in relevant cell types, such as hypothyroidism with T cells and thyroid, coronary artery disease with cardiomyocytes, hypertension and lipids with adipose tissue, and heart problems with heart ventricle and muscle cells. We replicated gene module associations with cardiovascular and autoimmune diseases in the Electronic Medical Records and Genomics (eMERGE) network. We also performed a CRISPR-screen to analyze lipid regulation in HepG2 cells and observed more consistent trait associations with modules than we observe with individual genes. Compared to a single-gene approach, our module-based method also better predicted FDA-approved drug-disease links by capturing tissue-specific pathophysiological mechanisms linked with the mechanisms of action of drugs (e.g. niacin with cardiovascular traits via a known immune mechanism). Exploring the phenotype-module space also revealed stable trait clusters across different resolutions, including a complex branch involving lipids with cardiovascular, autoimmune, and neuropsychiatric disorders. We offer a novel gene module approach to enhance the understanding of complex diseases and their therapeutic modalities.

PrgmNr 2893 - Quantifying factors that affect polygenic risk score performance in the eMERGE dataset

View session detail

Author Block: **D. Hui**[1], B. Xiao[1], R. R. Freimuth[2], G. P. Jarvik[3], L. C. Kottyan[4], I. J. Kullo[2], N. A. Limdi[5], C. Liu[6], L. Yuan[7], B. Namjou[8], M. J. Roy-Puckelwartz[7], W-Q. Wei[9], S. S. Verma[1], D. Kim[1], M. D. Ritchie[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Mayo Clinic, Rochester, MN, [3]Univ Washington Med Ctr., Seattle, WA, [4]Cincinnati Children's Hosp. Med. Ctr., Cincinnati, OH, [5]Univ. of Alabama at Birmingham, Birmingham, AL, [6]Columbia Univ., New York City, NY, [7]Northwestern Univ., Chicago, IL, [8]Children's Hosp. Med. Ctr., Cincinnati, OH, [9]Vanderbilt Univ, Nashville, TN

Disclosure Block: **D. Hui:** None.

Polygenic risk scores (PRS) have poor portability across ancestries, and evidence suggests that for some traits e.g., those related to adiposity, this disparity is exacerbated by environmental or personal-level variables such as age. Efforts to quantify how performance of PRS is affected when the study individuals and genome-wide association study (GWAS) summary statistics are not of homogeneous ancestry and personal background will provide insight into factors that limit transferability of PRS across populations. PRS were calculated for body mass index (BMI) on individuals from the eMERGE cohort (N=77,473), which includes individuals of diverse ancestry and ages (17% non-European ancestry, 19% less than age 21). Stratified analyses were conducted in three ancestry groupings and four age categories (2 than the UKBB GWAS (.0156 vs .0107, $p=2.72x10^{-19}$) in African ancestry adults, when both analyses used their best performing LD panel. In European ancestry adults, the UKBB GWAS had 24% higher R^2 than the GIANT GWAS (.0590 vs .0474, $p=2.03x10^{-164}$) - in both analyses, the LD panel using all individuals in 1000 Genomes performed best. Model performance was similar between adults and teenagers - using the best performing summary statistics and LD panels for each run, we observed differences in R^2 of .0590 vs .0591 ($p=8.52x10^{-8}$), .0504 vs .0417 ($p=8.57x10^{-11}$), and .0156 vs .0324 ($p=9.62x10^{-42}$) for European, all, and African ancestry individuals, respectively. Performance greatly degraded in children (R^2s of 0.0293, 0.0180, 0.0102 in European, all, and African ancestry individuals, respectively). In conclusion, we observed that 1) even small amounts of non-European ancestry individuals in the GWAS where effect estimates are derived greatly improved prediction performance in African ancestry individuals; 2) all of 1000 Genomes LD reference panels outperformed solely European ancestry LD panels even when both summary statistics and test individuals were of European ancestry; and 3) model performance was similar between adults and teenagers but degraded greatly for children.

PrgmNr 2894 - Rare variant GWAS of African American and Hispanic patients using WGS data from the TOPMed program on the NHLBI BioData Catalyst

View session detail

Author Block: **S. Gilhool**[1], F. D. Mentch[1], E. G. Burchard[2,3], A. C. Mak[2], L. K. Williams[4], H. Gui[4], S. Xiao[4], H. Hakonarson[1,5], P. M. Sleiman[1,5]; [1]Ctr. for Applied Genomics, Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Dept. of Med., Univ. of California, San Francisco, San Francisco, CA, [3]Dept. of Bioengineering and Therapeutic Sci., Univ. of California, San Francisco, San Francisco, CA, [4]Ctr. for Individualized and Genomic Med. Res., Dept. of Internal Med., Henry Ford Hlth.System, Detroit, MI, [5]Dept. of Pediatrics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA
Disclosure Block: **S. Gilhool:** None.
Asthma is the most common chronic respiratory disease, affecting over 300 million people worldwide. In the United States alone, asthma creates a financial burden of over $50 billion annually. This burden is not distributed equally; asthma is more prevalent and severe among African Americans and Hispanics, as compared to Caucasians. Asthma is known to be caused by a combination of environmental and genetic factors, but our knowledge of the genetic causes of asthma is still incomplete.

A small number of common genetic variants have been identified through genome wide association studies (GWAS), but these explain only a small proportion of asthma heritability. Furthermore, many GWASs have focused on Caucasian cohorts, leaving variants specific to other populations largely unknown.

In this project, we explore new ground in the genetic landscape of asthma by 1) probing rare variants, which could have larger effect sizes and may be population-specific, and 2) analyzing a cohort of traditionally under-studied populations. We leverage WGS data from the Asthma Translational Genomic Collaborative (ATGC) in the NHLBI Tran-Omics for Precision Medicine (TOPMed) program. In total, we analyze over 15,000 African American and racially admixed Hispanic patients.

The analysis is performed on NHLBI BioData Catalyst powered by Terra. The NHLBI BioData Catalyst facilitates this research by allowing TOPMed data to be easily imported to a scalable, high-performance cloud workspace, and to be analyzed using custom and community-developed workflows.

Hail is used to perform quality control on variant calls and filter for rare variants at two different frequency thresholds (MAF Our approach is focused on identifying novel genes harboring rare variants which are associated with asthma in under-studied populations. We anticipate our results will improve our understanding of the genetics of asthma, contribute to efforts to precisely tailor treatments to each patient's unique genetic background, and ultimately help reduce healthcare disparities.

PrgmNr 2895 - Rare, functional variants amongst 180,256 exomes from the UK Biobank influence mitochondrial copy number

View session detail

Author Block: **V. Pillalamarri**[1,2], C. Say[3], S. Y. Yang[1,2], C. E. Newcomb[1], N. Pankratz[4], D. E. Arking[1]; [1]McKusick-Nathans Dept. of Genetic Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [2]Predoctoral Program in Human Genetics, Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [3]Johns Hopkins Univ., Baltimore, MD, [4]Dept. of Lab. Med. and Pathology, Univ. of Minnesota Med. Sch., Minneapolis, MN

Disclosure Block: **V. Pillalamarri:** None.

Over 1.45 billion years since the endosymbiotic origin of mitochondria, mitochondrial function has evolved to become obligatory for eukaryotic cellular function and survival; mitochondria play vital roles in oxidative phosphorylation (oxphos), apoptosis and metabolism, and defects in human mitochondrial genomes (e.g. in genes encoding core subunits of oxphos) lead to disease. A portion of intercellular and interindividual variation in the number of mitochondrial genome copies (mtDNA-CN) reflects mitochondrial function and has recently been shown to be a heritable component under nuclear genetic control via common polygenic variation. We therefore hypothesized that nuclear genes would harbor an excess of rare, functional variants in aggregate that could influence variation in mtDNA-CN. Amongst 180,256 exomes from the UK Biobank cohort, we tested rare variants below one percent allele frequency, both at the level of single variants and through aggregate burden, SKAT and SMMAT analyses, for association with computed mtDNA-CN adjusted for cell counts and other relevant covariates. A survey across 16 variant sets composed of coding sequence and loss-of-function mutations at varying rare allele frequencies identified eight genes at Bonferroni significance and 21 genes at a false discovery rate TWNK ($p=1.03e^{-12}$) and mitochondrial transcription factor *TFAM* ($p=1.23e^{-5}$). Several novel gene associations included the essential iron transporter Mitoferrin-1 *SLC25A37* ($p=9.28e^{-7}$), *C20orf144* ($p=7.68e^{-7}$), and a single variant in the infertility gene *SPAG4* ($p=4.5e^{-15}$). The most significant association was with *SAMHD1* ($p=5.49e^{-30}$, q-value=$7.55e^{-25}$), a cytoplasmic host restriction factor involved with viral defense response and involved in the mitochondrial nucleotide salvage pathway. Using leave-one-out analyses, we pinpointed three functional variants driving the *SAMHD1* signal, the most explanatory being a single missense mutation at hg38:chr20:36,893,060.C>T ($p=1.4e^{-19}$, beta=+0.7, se=0.08 SD increase in mtDNA-CN) which falls within the critical region of Aicardi-GoutiÃ¨res syndrome 5 (AGS5) involved with childhood encephalopathy and chronic interferon-mediated inflammatory response. We hypothesize this missense variant within the AGS5 region disrupts the alternate role of SAMHD1 in resection of single-stranded DNA at stalled replication forks. This might lead to a buildup of cytosolic ssDNA, triggering an interferon-mediated inflammatory response via the cGAS-STING pathway and lead to an increase in mtDNA-CN.

PrgmNr 2896 - Relationship between autozygosity and complex traits in the Amish

View session detail

Author Block: **M. Lynch**[1], K. A. Maloney[2], H. Xu[3], A. R. Shuldiner[4], B. D. Mitchell[5]; [1]Univ. of Maryland Baltimore, Baltimore, MD, [2]Univ Maryland, Baltimore, MD, [3]Baltimore, MD, [4]Regeneron Genetics Ctr., Tarrytown, NY, [5]Univ Maryland, Baltimore, Baltimore, MD

Disclosure Block: **M. Lynch:** None.

Autozygosity, which measures portions of the genome that are homozygous by descent, has been associated with variation in traits of biomedical importance impacting evolutionary fitness in some populations. Estimates of autozygosity can be made from runs of homozygosity (ROHs) that arise when identical-by-descent (IBD) haplotypes are inherited from each parent. In population isolates with a small set of common founders and lack of mating with outside members, autozygosity measurements are elevated and members of such populations often have an increased burden of genetic disease due to increased homozygosity of rare, recessive variants. In this study, we examined the relationship between genome-wide autozygosity and complex traits in 7221 Old Order Amish individuals residing in Lancaster County, PA genotyped on a high-density genome wide array (Illumina Global Screening Array).

We obtained genome-wide estimates of individual autozygosity using the IBD tool implemented in Plink and calculated F_{ROH}, the proportion of the autosomal genome in runs of homozygosity above a specified length. The Amish are a relatively recent founder population, having immigrated from Europe to Lancaster 14-15 generations ago. In this population, the average length of an ROH segment was 5700 KB and the average number of segments per individual was 20, spanning ~3.7% of the genome. Measurements of F_{ROH} were then used as the primary predictor of various traits of interest in association analysis. We analyzed 72 traits including basic anthropometrics, blood pressure, fasting blood lipids, glucose, insulin, HbA1c, basic medical blood chemistry measurements, and medical histories. We did not identify any associations that withstood Bonferroni-correction, but the lead association ($p = 0.003$) with genome-wide F_{ROH} was with EKG QT interval. Although no traits were associated with autozygosity at the genome-wide level, we are currently estimating F_{ROH} regionally for incremental windows across the genome for use in autozygosity mapping to identify novel recessive trait associations.

PrgmNr 2897 - Single cell enrichment of expression and splicing QTL target genes mapped to GWAS loci identifies causal genes and pathogenic cell types for glaucoma and intraocular pressure

View session detail

Author Block: **A. R. Hamel**[1,2,3], J. Wang[1], J. M. Rouhana[1], G. Eraslan[4], F. Aguet[5], A. McAdams[1,6], W. Yan[6], A. Monavarfeshani[6], T. van Zyl[7,8], J. R. Sanes[6], J. L. Wiggs[1,2,3], A. V. Segre[1,2,3], International Glaucoma Genetics Consortium (IGGC); [1]Ocular Genomics Inst. and Dept. of Ophthalmology, Massachusetts Eye and Ear, Boston, MA, [2]Dept. of Ophthalmology, Harvard Med. Sch., Boston, MA, [3]Med. and Population Genetics Program, Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Klarman Cell Observatory, Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Cancer Program, Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Dept. of Molecular and Cellular Biology and Ctr. for Brain Sci., Harvard Univ., Cambridge, MA, [7]Regeneron Pharmaceuticals, Tarrytown, NY, [8]Dept. of Ophthalmology and Visual Sci., Yale Sch. of Med., New Haven, CT

Disclosure Block: **A.R. Hamel:** None.

Primary open-angle glaucoma (POAG), characterized by progressive optic neuropathy, is a leading cause of blindness worldwide, for which there are no cures and the pathophysiology is not well understood. Elevated intraocular pressure (IOP) is a major risk factor. A recent multi-ethnic genome-wide association study (GWAS) meta-analysis of 34,179 POAG cases and 349,321 controls identified 127 loci associated with POAG risk, and an IOP GWAS meta-analysis of 139,555 individuals identified 133 associated variants. However, the implicated causal genes and cell types are not known, as the majority of variants lie in noncoding, multi-genic regions, and few measurements of regulatory effects at single cell resolution are available. We developed a novel method, ECLIPSER (Enrichment of Causal Loci and Identification of Pathogenic cells in Single Cell Expression and Regulation data) that tests whether the expression of genes mapped to GWAS loci based on expression quantitative trait loci (eQTLs) and splicing QTLs (sQTLs) for a given complex disease or trait, is enriched in specific cell types. We used colocalization analysis or linkage disequilibrium between e/sQTLs and GWAS loci to map genes to loci. For each trait, tissue, and cell type combination, ECLIPSER scores GWAS loci by the fraction of cell type-specific genes (fold-change>1.4, FDRcis-eQTLs or sQTLs from 49 tissues in GTEx v8 or peripheral retina (EyeGEx), found â¥1 e/sQTLs likely to share a causal effect with 58% of GWAS loci (Posterior Prob>0.5), with an average of 4 eGenes and 2 sGenes per locus. These genes were enriched in cell-cell adhesion, ribonucleotide metabolism, negative regulation of cytokine production, and neuron differentiation. We next applied ECLIPSER to these POAG and IOP GWAS locus sets, using single nucleus expression data from the anterior segment of the eye, retina, and optic nerve head. We found significant enrichment for POAG and IOP in specific cell types, including Beam cells of the trabecular meshwork (TM) and vascular endothelial cells in the outflow pathway, implicated in elevated IOP. Sixteen known and new genes in the TM Beam cells were proposed to contribute to POAG risk, e.g. an sQTL acting on *ANTXR1*. This work proposes cell type-specific causal regulatory mechanisms and genes, and pathogenic cell types for glaucoma that could help guide novel therapeutic design. Our new method can be applied to other complex diseases.

PrgmNr 2898 - Top-LD: a tool to explore linkage disequilibrium using TOPMed whole genome sequence data

View session detail

Author Block: **L. Huang**[1], J. D. Rosen[2], Q. Sun[2], J. Chen[2], Y. Zhou[3,4], S. S. Rich[5], M. P. Conomos[6], A. Stilp[6], C. McHugh[6], J. I. Rotter[7], R. J. F. Loos[8], E. Kenny[8], A. W. Manichaikul[5], A. Correa[9], C. Kooperberg[4], L. M. Raffield[10], A. P. Reiner[3,4], J. Roach[11], NHLBI Trans-Omics for Precision Medicine (TOPMed)Consortium, TOPMed Analysis Working Group, P. L. Auer[12], Y. Li[2,10,13]; [1]Curriculum in Bioinformatics and Computational Biology, UNC-CH, Chapel Hill, NC, [2]Dept. of Biostatistics, UNC-CH, Chapel Hill, NC, [3]Dept. of Epidemiology, Univ. of Washington, Seattle, WA, [4]Fred Hutchinson Cancer Res. Ctr., Univ. of Washington, Seattle, WA, [5]Dept. of Publ. Hlth.Sci., Univ. of Virginia Sch. of Med., Charlottesville, VA, [6]Univ. of Washington, Seattle, WA, [7]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [8]The Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York City, NY, [9]Dept. of Med., Univ. of Mississippi Med. Ctr., Jackson, MS, [10]Dept. of Genetics, UNC-CH, Chapel Hill, NC, [11]Res. Computing, UNC-CH, Chapel Hill, NC, [12]Univ. of Wisconsin Milwaukee, Milwaukee, WI, [13]Dept. of Computer Sci., UNC-CH, Chapel Hill, NC

Disclosure Block: **L. Huang:** None.

Current publicly available tools that allow rapid exploration of linkage disequilibrium (LD) between markers (e.g., HaploReg and LDlink) are based on low-coverage (2-6X) whole-genome sequence (WGS) data from 1,092 individuals in the 1000 Genomes Project. Here, we present Top-LD, an online tool to explore LD inferred using Trans Omics for Precision Medicine (TOPMed) high-coverage (~30X) WGS data from 15,578 individuals, which include 13,160 individuals of European ancestry, 1,335 individuals of African ancestry, 844 individuals of East Asian ancestry, and 239 individuals of South Asian ancestry. Individuals were selected as those having > 90% ancestry from one of these populations, as estimated from the TOPMed data using RFMix. Top-LD provides a significant upgrade compared to current LD tools as the TOPMed WGS data provide a much more comprehensive representation of genetic variation in these four populations compared to the 1000 Genomes Project. For example, Top-LD encompasses LD information on 134.4, 43.9, and 27.3 million variants with minor allele frequency S1PR3 that is in near perfect LD (Rsq=0.99) with a GWAS sentinel SNP (rs28450540) for circulating monocyte counts. We will present this and other examples of information from Top-LD that have aided fine-mapping of known loci. Top-LD is freely available at http://topld.genetics.unc.edu/topld.

PrgmNr 2899 - Trans-ethnic Transcriptome-wide Association Study and fine-mapping analysis in 3.4 million individuals shed lights on the genetic architecture of alcohol and smoking addiction

View session detail

Author Block: **F. Chen**[1], X. Wang[1], G. Saunders[2], the GWAS & Sequencing Consortium of Alcohol and Nicotine use, Trans-Omics for Precision Medicine (TOPMed) Program; [1]Penn State Coll. of Med., Hershey, PA, [2]Univ. of Minnesota, Minneapolis, MN
Disclosure Block: **F. Chen:** None.
Cigarette smoking is a well-established major heritable risk factor for human diseases. And alcohol consumption has been increasingly associated with the development of chronic diseases and other serious problems, such as heart disease, stroke, and cancer. The availability of large datasets in recent years enabled a breakthrough in the genetics of smoking and alcohol addiction, with more than 400 loci discovered to date (most are with smoking). The identified loci numbers keep growing thanks to the rapidly increasing sample size and involvement of ethnically diverse populations. But it remains challenging to map these non-coding variants to their target genes and translate their biological and clinical relevance. To address this gap, we developed a novel trans-ethnic TWAS approach, TESLA (trans-ethnic transcriptome-wide association study approach using an optimal linear combination of association statistics), that has provably optimal power to integrate trans-ethnic GWAS with eQTL data from a possibly unmatched ancestry. TESLA uses a mixed effect meta-regression model to model ancestry-specific effect across different studies in meta-analysis. We showed using simulation that TESLA substantially outperforms other strategies, including TWAS using fixed effect meta-analysis results and TWAS using only studies from matched ancestries. Using this approach, we aggregated trans-ethnic GWAS and whole-genome sequencing data from GSCAN2 (total N = 3.4 million) and GTEx v8 eQTL data in 49 tissues to further empower gene discovery for tobacco and alcohol use behaviors. We identified 2,652 genes; among them, 148 are novel genes that are outside 1 million basepair window of GWAS sentinel variants. These results provide us much more susceptibility genes for alcohol and smoking addiction than hitherto. For example, besides the well-studied *CHRNA3, CHRNA5, IREB2* and *PSMA4* genes, we also identified novel genes such as *HMGN3* and *LATS1* associated with CigDay (Cigarettes per Day) in multiple tissues. Consistent with previous studies, our results also showed a general lack of tissue specificity across all phenotypes where the most significant genes in each tissue are often ranked the top across many tissues. Further, we performed a fine-mapping analysis in these risk TWAS loci to prioritize putative causal genes. The 90%-credible set of 778 loci were narrowed down to a single gene. These results could help us develop a deeper understanding and broad vision of the genetic architecture of smoking and alcohol use behaviors.

PrgmNr 2900 - Whole exome sequencing analysis identified five novel genes associated with Osteoarthritis

View session detail

Author Block: **S. Werdyani**[1], Q. Li[2], M. Liu[1], A. Dohey[3], D. ORielly[3], P. Rahman[4], G. Zhai[1]; [1]Div. of BioMed. Sci. (Genetics), Faculty of Med., Mem. Univ. of Newfoundland, St. John's, NL, Canada, [2]Princess Margaret Cancer Ctr., Univ. Hlth.Network, Univ. of Toronto, Toronto, ON, Canada, [3]Molecular Genetics and Translational Genomics Lab., Eastern Hlth., St. John's, NL, Canada, [4]Discipline of Med., Faculty of Med., Mem. Univ. of Newfoundland, St. John's, NL, Canada
Disclosure Block: **S. Werdyani:** None.
Background: Osteoarthritis (OA) is the most common chronic progressive joint condition and one of the ten most disabling diseases worldwide. Pathogenesis of OA remains elusive but is thought to be caused by interaction of genetic and environmental factors. This study was to identify genes for knee and hip OA by whole exome sequencing (WES) analysis from the well-established Newfoundland Osteoarthritis Study (NFOAS).
Methods: Study participants were total knee or hip replacement patients due to primary OA who were recruited to the NFOAS before 2017 in St. Johnâ□□s, Canada. Patientsâ□□ blood DNA was extracted and sequenced by the IlluminaÂ® NovaSeq 6000 at the Translational Genomics Laboratory of Memorial University. GATK Best Practices pipeline was followed to align the WES raw paired end reads to the GRCh37 human reference genome, produce master binary alignment map (BAM) files, and create variants calling format (VCF) files. The resulted VCF files were used for the functional variant annotation using ANNOVAR software. Quality control (QC) filtering was set to exclude genetic variants with â□¤ 20X overall depth of coverage and â□¤ 5X for alternative allele. Further, variants having minor allele frequency (MAF) â□¤ 0.01 and â□¥ 0.1 in the general population based on the 1000 genomes project, ExAC, and gnomAD databases were excluded. Finally, only variants present in â□¥ 80% of OA patients were considered as potential OA associated genetic variants.
Results: A number of 144 knee and 56 hip OA patients from the NFOAS (mean age 62.83Â±7.61 years, and 51.5% of them were females) were included in the study. A total of 526,459 genetic variants were identified in the cohort, after the QC, 92 of them were identified to be associated with OA. These genetic variants presented in â□¥ 80% of the OA patients with a MAF of ~ 0.49, for both knee and hip OA. The MAFs of these variants in all public available databases were â□¤ 0.06. Ten exonic nonsynonymous SNV located in *IGSF3*, *ZNF717*, *PRSS1*, *AQP7*, and *ESRRA* genes were estimated to have significant damaging effects on the related proteinsâ□□ structure and function by > 4 functional and pathogenic prediction tools. These genes play a central role in metabolism of water-soluble vitamins and cofactors, aquaporin-mediated transport, glucagon signaling in metabolic regulation, regulation of lipolysis in adipocytes, and the extracellular matrix degradation, but have not been reported in previous OA GWAS studies.
Conclusion: Our data showed five novel genes to be associated with OA. While confirmation is required, the findings provided new insights into better understanding of knee and hip OA pathogenesis and hold promising as druggable targets for developing OA therapies.

PrgmNr 2901 - A comprehensive functional screen detects multiple polymorphic RET enhancers affecting Hirschsprung disease risk

View session detail

Author Block: **L. Fries**[1], K. M. Karasaki[2], A. Kapoor[3], S. Chatterjee[4]; [1]Ctr. for Human Genetics and Genomics, NYU Sch. of Med., New York, NY, [2]Dept. of Biology, Johns Hopkins Univ., Baltimore, MD, [3]UTHSC, Houston, TX, [4]NYU Sch. of Med., New York, NY
Disclosure Block: **L. Fries:** None.
Hirschsprung disease (HSCR), a congenital disorder characterized by enteric aganglionosis, arises from both rare coding and common enhancer variants in the receptor tyrosine kinase gene RET. Though both genome wide association studies (GWAS) and whole genome sequencing have demonstrated the association of multiple non-coding variants at the locus, there hasnâ□□t been a comprehensive functional screen to determine which of these variants disrupt cis-regulatory elements (CREs) and affect *RET* function.Our GWAS on 220 European ancestry trios has deciphered 38 common (minor allele frequency, MAF â□¥10%), non-coding HSCR associated single nucleotide polymorphisms (SNPs) at the RET locus. Functional screens on 8 of these SNPs led to the discovery of 3 independent SNPs, residing within enhancers, that disrupt binding of the transcription factors (TF) RARÎ², GATA2 and SOX10 and affect the transcription of RET and multiple genes of the RET gene regulatory network (GRN).Using the human neuroblastoma cell line SK-N-SH we have now tested the remaining 30 HSCR associated common variants to demonstrate that 22 lie within CREs and of which 7 have sequence variants with differential enhancer and RET gene expression activity in addition to the 3 identified before. We demonstrate that the risk of HSCR is significant for two haplotypes, GCAGGTTGGT (OR 12.2, 95% CI: 5.97-24.93, P=7.02x10-12) and CTGAGTTGGT (OR 7.2, 95% CI: 3.26-15.91, P=1.02x10-6). Two of the new CREs bind transcription factor PAX3, known to affect neural crest cell migration and ENS differentiation. Reducing PAX3 expression using siRNA decreases RET gene expression as well as that of SOX10, the major TF in the RET GRN. Deleting these SNP containing RET CREs using CRISPR/Cas9 genome editing reduces RET expression by only 20-30% but is insufficient to perturb gene expression for other GRN genes as those effects are evident only when RET gene expression falls below 50% of its wildtype level. Therefore, in HSCR, significant reductions of gene expression at RET and its GRN to affect ENS development only occurs in individuals with multiple CRE variants in combination with variants in other GRN genes.

PrgmNr 2902 - Association of DNA Methylation with Mixed Substance Use in an HIV-positive Cohort

View session detail

Author Block: **M. Lin**, A. Justice, K. Xu; Yale Univ., New Haven, CT
Disclosure Block: **M. Lin:** None.
Background Substance use is common in people with HIV and results in worsening HIV progression even under antiretroviral therapy. People may alternately use different substances at different period, referred as mixed substance use (MSU). Recent studies link epigenetic mechanisms with individual substance use such as tobacco smoking, alcohol, or cocaine. However, no study of DNA methylation and MSU has been reported. This study aimed to test an association of DNA methylation in circulating blood of HIV-positive patients who used different level of substances from a longitudinal cohort, Veteran Aging Cohort Study (VACS).
Methods This study included 561 male HIV-positive African Americans. Confirmatory factor analysis was conducted to extract a second-order factor of substance use, accounting for shared variance between tobacco, cannabis, alcohol, cocaine, and heroin use and across three follow-up timepoints. Severity of each substance used was assessed by self-reported frequency of use for each substance at each time point. DNA methylation was profiled with Illumina HumanMethylation EPIC Beadchip. The methylation profiles were filtered and functional normalized with *minfi* package. Singular value decomposition was conducted to remove noise from data. Differential methylation CpG position (DMP) and region (DMR) were analyzed and followed by Gene Ontology and pathway analysis.
Results Confirmatory factor analysis revealed three categories of MSU: Severe (N=166, age=47.83Â±5.94), Moderate (N=305, age=48.06Â±7.77) and Mild (N=90, age=49.81Â±9.45). In comparing Severe and Mild MSU groups, we identified 169 DMPs and 10 DMRs (False Discovery Rate, FDRAHRR, ELMSAN1, SEPT9, MATR3, SIL1, NSMCE1, PRR15, C17orf97, and two long non-coding genes *AC068134.5* and *AC108004.3*. Compared to mild MSU group, individuals with severe MSU showed 69 hypermethylated CpG sites and 100 hypomethylated CpG sites. The genes harboring 169 CpG sites were enriched on inflammatory bowel disease pathway in KEGG pathway database (P BMERB1 gene (FDR = 0.041). No differentially methylated CpG site was identified in comparing Severe to Moderate groups.
Conclusion Our results indicate that mixed substance use is associated with altered DNA methylation. The enrichment of MSU-associated DNA methylation on biological pathways may enhance our understanding of the underlying mechanisms of HIV and its comorbidity.

PrgmNr 2903 - Cis-regulatory hubs constitute a powerful model to understand the impact of 3d organization in schizophrenia

View session detail

Author Block: **L. Mangnier**[1], C. Joly-Beauparlant[2], A. Droit[2], S. Bilodeau[3], A. Bureau[4]; [1]Laval Univ., Medecine Faculty, Quebec, QC, Canada, [2]Ctr. de Recherche du CHU de QuÃ©bec - UniversitÃ© Laval, Quebec, QC, Canada, [3]Ctr. de recherche du CHU de QuÃ©bec â UniversitÃ© Laval, Axe Oncologie, Quebec, QC, Canada, [4]Univ Laval, QuÃ©bec, QC, Canada
Disclosure Block: **L. Mangnier:** None.
The cis-regulatory modules (CRMs) are noncoding regulatory regions, playing a crucial role in the regulation of transcription and the emergence of complex phenotypes. Recent single-cell multi-way analyses show that several CRMs and genes locally co-interact through 3D contacts, building hubs. Despite the importance of CRM hubs in gene regulation, their precise implication in complex diseases such as schizophrenia remains unclear. In the present study, we model cis-regulatory hubs (CRHs), using available Hi-C data in neurons derived from induced pluripotent stem cells and the activity-by-contact model linking active enhancers to promoters. Comparing CRHs to either equivalent tissue-specific or non-tissue-specific structures, we showed that they constitute a relevant organization for schizophrenia. Firstly, we defined CRHs as active structures, associated with gene activity. Then, we assessed the relevance of CRHs in schizophrenia using H-Magma. Considering the noncoding SNPs in 3D contact with genes, we found an enrichment in schizophrenia-associated genes within CRHs compared to genes outside (OR=1.81). Next, using the linkage disequilibrium score regression we showed that a larger portion of schizophrenia heritability is explained by CRHs than non-tissue-specific elements, with enrichment of 3 against 0.43 on average. In addition, we also observed up to 11-fold enrichment in schizophrenia heritability compared to equivalent tissue-specific elements. This result is supported by schizophrenia-associated SNP enrichment (OR=1.29) and tissue-relevant GO pathway analysis. Our results demonstrate that CRHs in neurons constitute a useful model for understanding the 3D organization between CRMs and genes involved in the emergence of complex phenotypes such as schizophrenia.

PrgmNr 2904 - Developmental regulation of neuronal gene expression by Elongator complex protein 1 dosage

View session detail

Author Block: **E. Morini**[1], G. Dadi[1], E. Logan[1], M. Salani[1], A. J. Krauson[1], A. Chekuri[1], Y-T. Chen[2], A. Ragavendran[1], P. Chakravarty[3], S. Erdin[1], A. Stortchevoi[1], J. Q. Svejstrup[3], M. E. Talkowski[1], S. A. Slaugenhaupt[4]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]Dept. of Life Sci. and Inst. of Genome Sci., Natl. Yang Ming Chiao Tung Univ., Taipei, Taiwan, [3]Bioinformatics and Biostatistics, The Francis Crick Inst., London, United Kingdom, [4]Massachusetts Gen. Hosp. Res. Inst., Boston, MA
Disclosure Block: **E. Morini:** None.
Elongator is a highly conserved protein complex required for transcriptional elongation, intracellular transport and translation. Elongator complex protein 1 (ELP1) is the scaffolding protein of Elongator and is essential for its assembly and stability. Familial dysautonomia (FD), a hereditary sensory and autonomic neuropathy, is caused by a mutation in *ELP1* that leads to a tissue-specific reduction of ELP1 protein. Our work to generate a phenotypic mouse model for FD led to the discovery that homozygous deletion of the mouse *Elp1* gene leads to embryonic lethality prior to mid-gestation. Given that FD is caused by a reduction, not loss, of ELP1, we generated two new mouse models by introducing different copy numbers of the human FD *ELP1* transgene into the *Elp1* knockout mouse (*Elp1*$^{-/-}$) and observed that human *ELP1*expression rescues embryonic development in a dose dependent manner. We then conducted a comprehensive transcriptome analysis in mouse embryos to identify genes and pathways whose expression correlates with the amount of *ELP1*. We found that *ELP1* is essential for the expression of genes responsible for the formation and development of the nervous system. Further, gene length analysis of the differentially expressed genes showed that the loss of *Elp1* mainly impacts the expression of long genes and that by gradually restoring Elongator their expression is progressively rescued. Finally, through evaluation of co-expression modules, we identified gene sets with unique expression patterns that depended on *ELP1* expression. Overall, this study highlights the crucial role of *ELP1* during early embryonic neuronal development and reveals gene networks and biological pathways that are regulated by Elongator.

PrgmNr 2905 - Differential methylation signatures in atypical parkinsonism syndromes

View session detail

Author Block: **P. Reho**[1], T. G. Beach[2], G. E. Serrano[2], K. L. Newell[3], S. Solaiman[1], H. P. Easwaran[4], J. C. Troncoso[5], O. Pletnikova[6], S. W. Scholz[1,7]; [1]Neurodegenerative Diseases Res. Unit, Natl. Inst. of Neurological Disorders and Stroke, Bethesda, MD, [2]Banner Sun Hlth.Res. Inst., Sun City, AZ, [3]Dept. of Pathology and Lab. Med., Indiana Univ. Sch. of Med., Indianapolis, IN, [4]Dept. of Oncology, The Johns Hopkins Sch. of Med., The Sidney Kimmel Comprehensive Cancer Ctr., Baltimore, MD, [5]Dept. of Pathology (Neuropathology), Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [6]Dept. of Pathology and Anatomical Sci., Jacobs Sch. of Med. and BioMed. Sci., Univ. of Buffalo, Buffalo, NY, [7]Dept. of Neurology, Johns Hopkins Univ. Sch. of Med., Baltimore, MD

Disclosure Block: **P. Reho:** None.

Atypical parkinsonism (AP) refers to a group of heterogeneous, neurodegenerative disorders that include multiple system atrophy (MSA), Lewy body dementia (LBD), progressive supranuclear palsy (PSP), and corticobasal degeneration. Although AP syndromes are clinico-pathologically distinct entities, the clinical presentations commonly overlap and lead to delayed diagnosis. The causes of these disorders are poorly understood, and treatments are limited to supportive care. Despite recent discoveries of genetic risk factors, the role of epigenetic mechanisms is still largely unknown. Cytosine methylation changes DNA accessibility to the transcriptional machinery complex, finely modulating gene expression. These epigenetic events play essential roles during development and can vary across the lifespan in response to lifestyle, environmental factors, as well as genetic variation. We performed an epigenome-wide association study of 363 pathologically defined atypical parkinsonian patients and 131 neurologically healthy controls. We analyzed 230 LDB, 49 MSA and 84 PSP patients. The mean age of death was 78 years for control and LBD subjects, while 73 years and 65 years for PSP and MSA patient respectively. Males and females were equally distributed in our study cohort. We hybridized 450 ng of bisulfite-converted DNA, extracted from cerebellar cortex, to Infinium MethylationEPIC Beadchips (Illumina). The raw data were preprocessed in Genome Studio (Illumina) and exported to the R package â□□minfiâ□□ for further analysis. Low-quality samples (p-value > 0.01), probes located on X and Y chromosomes or that had SNPs at any CpG sites, and probes that failed in at least one individual were excluded from the study. A total of 772,295 probes were available for the analysis. We detected 6,314 differentially methylated probes (adj. p-value -8) located at the transcription start site of 4,571 genes: 2,813 (62%) were hypermethylated and 1,758 (38%) were hypomethylated in AP patients. Differential DNA methylations mainly affected biological processes implicated in neuron differentiation (GO:0030182, adj. p-value $7.55x10^{-15}$) and neurogenesis (GO:0022008, adj. p-value $2.58x10^{-15}$). Notably, differential methylation regions analysis detected 21 modulated probes at the promoter of *ZFP57* gene (Fisher $6.37x10^{-9}$). Preliminary results from our study show a cerebellar DNA hypermethylation signature in AP patients compared to controls, suggesting that epigenetic regulation plays an important role in AP pathogenesis. These insights may pave the way for the identification of new molecular pathways that could be targeted for the development of novel therapeutics.

PrgmNr 2906 - Dissecting multiple-signal GWAS loci and their regulatory roles in Alzheimerâ□□s Disease

View session detail

Author Block: **M. Ionita**, P. P. Kuksa, C. Chen, Y. Leung, L-S. Wang; Penn Neurodegeneration Genomics Ctr., Perelman Sch. of Med. at the Univ. of Pennsylvania, Philadelphia, PA
Disclosure Block: **M. Ionita:** None.
Introduction:
Recent large-scale genome-wide association studies (GWAS) have identified many genomic loci associated with Alzheimerâ□□s disease (AD), most of which are non-coding and pose the additional challenges of identifying underlying non-coding mechanisms and target genes. Robust methods are needed to perform a systematic screening of all genome-wide significant loci from AD GWAS (Kunkle et al, 2019; N=94,437) for multiple (primary and secondary) colocalized variants (genetic signals).
Method:
Using a custom version (https://github.com/matei-ionita/INFERNO_eCAVIAR) of the eCAVIAR method (Hormozdiari et al, 2016), we first identified genome-wide significant GWAS loci with gene-regulatory evidence based on tissue-specific GTEx expression quantitative trait loci (eQTL) data. Colocalized loci were defined as loci with cumulative region-level colocalization probability > 0.5 across tested single- and multiple- variant configurations. For these colocalized gene-regulatory loci, we then performed a multiple-signal analysis to identify all primary and secondary signals with corresponding target genes across 48 tissues.
Results:
We identified 89 non-HLA colocalized loci with 244 target genes across all tissues. We found that 48% of the colocalized AD GWAS loci had evidence for multiple signals in at least one tissue. 14% (1904/13583) of the tested locus-gene-tissue combinations satisfied the colocalization criteria and 14.2% (271/1904) of those had multiple underlying signals. Intriguingly, most colocalized variants (83%) were different from the tag GWAS variant.
We found that colocalized variants were primarily located within non-coding genomic regions, with 50% mRNA intronic, 34% long non-coding RNA, 6.9% UTR signals, with an average distance of 391kb between a colocalized variant and the target gene.
On average, we found 4.18Â±3.42 regulatory variants per identified locus across all tissues, suggesting multiple underlying tissue-specific mechanisms at each GWAS locus. Similarly, we observed tissue-specific effects on the target genes: each identified variant was found to be regulating on average 5.69 genes in 6.2 tissues.
When validated using matching tissue functional genomic data (FILER; Kuksa et al, 2021), the identified variants were located in open chromatin (for 78% of loci) and enhancer regions (for 68% of loci), confirming their active functional roles.
Altogether, our analysis of AD GWAS has identified both primary and secondary signals for the majority (55%) of target genes, including MTCH2, MADD, BIN1 in brain, digestive tissues and blood, recapitulating some of the reported molecular mechanisms for the disease.

PrgmNr 2907 - DNA methylation signature of *ASXL1* variants causing Bohring-Opitz syndrome (BOS)

View session detail

Author Block: **Z. Awamleh**[1], E. Charter-Diehl[1], S. Choufani[1], V. A. Arboleda[2], B. E. Russell[3], R. Weksberg[1]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of California Los Angeles, Los Angeles, CA, [3]Univ. of California, Los Angeles, Los Angeles, CA
Disclosure Block: **Z. Awamleh:** None.
The additional sex combs-like (ASXL) gene familyâ□□encoded by *ASXL1*, *ASXL2*, and *ASXL3*â□□is crucial for mammalian development through transcriptional regulation of the *HOX* gene cluster. Pathogenic variants in *ASXL* genes are associated with three phenotypically distinct neurodevelopmental syndromes. Our previous work has shown that syndromic conditions caused by pathogenic variants in epigenetic regulatory genes show consistent patterns of DNA methylation (DNAm) in peripheral blood: DNA methylation signatures. With evidence of the role of ASXL1/2 in chromatin modification, particularly deubiquitylation of lysine (K119) on histone 2A, we hypothesized that pathogenic *ASXL1* variants underlying Bohring-Opitz syndrome (BOS) have a unique DNAm signature. We profiled whole-blood DNAm for 17 *ASXL1* variants, 1 *ASXL2* variant, 7 *ASXL3* variants, and 40 sex- and age-matched typically developing individuals, using Illuminaâ□□s Infinium EPIC array. Using linear regression modelling, we identified 763 differentially methylated CpG sites (q 10%) in individuals with BOS (n=8). Differentially methylated sites overlap 323 unique genes, including *HOXA5* and *HOXB4*, supporting the functional relevance of DNAm signatures. A machine-learning classification model based on the DNAm signature classified a validation cohort of BOS individuals (n=6) with *ASXL1* variants and controls (n=100) correctly, demonstrating 100% sensitivity and specificity. We used the model to classify variants of uncertain significance in *ASXL1* (n=3) as pathogenic or benign; each classification was congruent with the patientsâ□□ clinical phenotypes. We used the machine-learning classification model to classify *ASXL2* and *ASXL3* variants. The *ASXL2* variant classified as pathogenic, whereas the *ASXL3* variants classified as benign. This suggests that *ASXL2* has a DNAm profile overlapping the *ASXL1* DNAm profile, whereas *ASXL3* has a unique profile. Next, we used the DNAm data to investigate DNAm age acceleration using Horvathâ□□s epigenetic clock. We observed acceleration in DNAm age in individuals with pathogenic *ASXL1* variants and the BOS phenotype, and the individual with the pathogenic *ASXL2* variant, but not in individuals with benign *ASXL1* variants or any *ASXL3* variants. Our DNA methylation data provide unique insights into *ASXL* syndrome pathophysiology. The DNAm signature we generated is gene- and syndrome-specific. It classifies all individuals with Sotos, Weaver, and Kabuki syndromes as benign, accurately differentiating pathogenic *ASXL1* variants causing BOS from other neurodevelopmental disorders caused by pathogenic variants in chromatin modifying genes.

PrgmNr 2908 - Epigenome-wide association study of circulating IgE levels identifies novel targets for asthma

View session detail

Author Block: **K. Recto**[1,2], P. Kachroo[3], T. Huan[1,2], G. Lee[1,2], D. Lee[1,2], J. Gereige[4], C. Yao[1,2], S-J. Hwang[1,2], R. Joehanes[1,2], G. O'Connor[4], S. T. Weiss[3], D. Levy[1,2], D. L. DeMeo[3]; [1]Framingham Heart Study, Framingham, MA, [2]Population Sci. Branch, Natl. Heart, Lung, and Blood Inst., Bethesda, MD, [3]Channing Div. of Network Med., Brigham and Womenâ□□s Hosp., Harvard Med. Sch., Boston, MA, [4]Pulmonary Ctr., Boston Univ. Sch. of Med., Boston, MA

Disclosure Block: **K. Recto:** None.

Measurement of circulating immunoglobulin E (IgE) concentration is a valuable aid in diagnosing and treating IgE-associated conditions, most notably asthma and allergic diseases. Identifying novel epigenetic signatures associated with serum IgE is critical to improve the understanding of the molecular mechanisms underlying IgE regulation. To this end, we performed an epigenome-wide association study (EWAS) to identify differential DNA methylation CpG sites associated with circulating IgE levels from 3471 participants in the Framingham Heart Study (FHS). We identified 490 statistically significant CpGs at a false discovery rate (FDR) less than 5%. We next validated our results using two independent cohorts: the Childhood Asthma Management Program (CAMP; n=798) and the Genetic Epidemiology of Asthma in Costa Rica Study (CRA; n=821). Of the 490 FDR-significant CpGs from FHS, 193 loci also passed an FDR less than 5% in both CAMP and CRA cohorts. Gene ontology analysis of the closest gene to the associated CpGs revealed that many of the genes were enriched in pathways related to transcription factor binding, asthma, and other immune system processes. We then analyzed the expression of genes associated with the 193 IgE-associated CpGs (i.e. expression quantitative trait methylation loci; eQTMs), which identified 61 *cis*-eQTMs at an FDR less than 5%. Lastly, using results of genome-wide association studies (GWAS) of IgE and IgE-related diseases, including asthma and allergy, we performed Mendelian Randomization (MR) for the 490 IgE-associated CpGs from FHS with methylation quantitative trait loci (mQTLs) as instrumental variables. MR analysis revealed 13 loci as nominally significant (pFCER1A, a known susceptibility locus for IgE levels. The majority of these associated sites have been previously associated with inflammation, asthma, and other immune functions. To date, there have been no previously published EWAS of IgE that integrate both eQTM and MR analyses. As such, our findings build upon prior knowledge of IgE regulation by providing a deeper understanding of the multidimensional inter-relations of DNA methylation, gene expression, and IgE levels. The IgE-associated loci that we identifiedâ□□particularly those implicated by eQTM and MR analysesâ□□can be explored as promising therapeutic targets for asthma and IgE-related diseases.

PrgmNr 2909 - Genetic effects on brain traits impact cell-type specific alternative splicing during human neurogenesis

View session detail

Author Block: **N. Aygun**[1], A. L. Elwell[1], D. Liang[2], M. Lafferty[2], K. E. Cheek[1], K. P. Courtney[1], J. Mory[1], E. Hadden-Ford[1], O. Krupa[1], L. de la Torre-Ubieta[3], D. H. Geschwind[4], M. I. Love[1], J. Stein[1]; [1]UNC Chapel Hill, Chapel Hill, NC, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Univ. of California Los Angeles, Los Angeles, CA, [4]Univ California los Angeles, Los Angeles, CA
Disclosure Block: **N. Aygun:** None.
Genetic regulation of splicing is an important risk factor for neuropsychiatric disorders that can be discovered via splicing quantitative trait locus (sQTL) analysis. sQTL analysis is generally performed in bulk post-mortem adult tissue. However, genetic risk loci are enriched in developmentally active regulatory elements, and the impact of risk variants may be masked by heterogeneity in bulk tissue. Here, we performed RNA-seq on a homogeneous population of primary human neural progenitor cells (Ndonor = 85) and their 8-week differentiated, virally labeled, and sorted neuronal progeny (Ndonor = 74), and genotyped donors of each cell line on a dense array (Illumina Omni 2.5+Exome) followed by imputation to a common reference panel (1000 Genomes).We detected sQTLs via testing association between splice junctions and genetic variants located within a Â± 200 kb window from the start and end of junctions per cell-type. We identified 4,568/3,870 intron excisions within 2,275/2,042 sGenes associated with 5,873/4,396 conditionally independent sSNPs-intron junction pairs in progenitors/neurons, respectively (at 5% false discovery rate for a hierarchical multiple testing correction). 79.3%/73.4% of conditionally independent sQTLs in progenitor/neuron were not discovered in fetal bulk brain sQTLs or in adult bulk sQTL data from GTEx using LD-based overlap. Applying co-localization analysis, we observed 29, 20, and 34 brain related trait GWAS loci in total that co-localized with specifically progenitor/neuron sQTLs and sQTLs present in both cell types, respectively. 111 cell-type specific sQTL trait associated loci-intron junction pairs were not fetal bulk cortical sQTLs. For instance, a progenitor specific sSNP (rs1222218, p-value = 5 x 10-9) regulating a novel alternative exon skipping event for *ARL14EP* gene playing a role in axonal development, was colocalized with schizophrenia (SCZ) index SNP (rs1765142). Lastly, applying transcriptome-wide association study (TWAS) approach, we detected the cis-heritable impact of 372/370 intron junctions in progenitor/neuron significantly correlated with at least one brain related-traits. As an example, splicing of an intron of *MRM2* gene encoding a mitochondrial rRNA methyltransferase was associated with increased risk for SCZ specifically in progenitor cells (TWAS-Z: 6.54), but it was not significantly cis-heritable within neuron, fetal bulk or adult bulk data. With this study, we propose a list of novel cell-type and temporal specific genetically altered alternative splicing in the human brain that can be candidate pathways to uncover cellular mechanisms underlying neuropsychiatric disease risk.

PrgmNr 2910 - Leveraging the Mendelian Disorders of the Epigenetic Machinery to Systematically Map Functional Epigenetic Variation

View session detail

Author Block: **L. Boukas**[1,2], T. R. Luperchio[1], K. D. Hansen[2], H. T. Bjornsson[1]; [1]Dept. of Genetic Med., Johns Hopkins Sch. of Med., Baltimore, MD, [2]Dept. of Biostatistics, Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD

Disclosure Block: **L. Boukas:** None.

A long-standing problem in epigenetics is the identification of specific 'epigenotypes' that *causally* mediate disease phenotypes via the alteration of transcriptional states. While for many diseases statistical associations have been detected, it is challenging to rule out the influence of confounders like environmental exposures, and to decide whether these epigenetic changes are a cause or consequence of the disease process.

We propose an approach to overcome these issues, by leveraging the Mendelian Disorders of the Epigenetic Machinery (MDEMs). In MDEMs, a coding variant disrupts an epigenetic regulator, leading to epigenetic abnormalities, which give rise to the phenotype, likely through a perturbation of the transcriptome. Although each MDEM has a different causative gene, at the phenotypic level MDEMs share common manifestations. We hypothesize that this phenotypic convergence results from shared epigenetic/transcriptomic alterations, and that identifying these will provide a catalog of loci/genes where epigenetic/transcriptomic variation is causally related to disease phenotypes. As proof-of-principle, we use mouse models of Kabuki syndromes type 1 and 2 (KS1 and 2; caused by haploinsufficiency of histone methyltransferases KMT2D and KDM6A, respectively), and Rubinstein Taybi type I (RT1; caused by haploinsufficiency of histone acetyltransferase CREBBP). We focus on the immune dysfunction exhibited by all three syndromes, which includes hypogammaglobulinemia and abnormal B cell development. We perform ATAC-seq and RNA-seq on sorted B cells from mutant mice, and from age- and sex-matched wild-type littermates from the same strain, thus controlling for both genetics and the environment.

We develop a new statistical approach for this kind of analysis, which we show provides us with increased power. We discover that 69% of promoters disrupted in KS1 are disrupted in KS2 also, with 67% of the shared KS1/2 promoters disrupted in RT1. This overlap is also present at enhancers, but to a lesser extent (51% shared between KS1/2, with 36% of those shared with RT1). We show that disruption of chromatin accessibility at promoters often leads to dysregulation of downstream gene expression, and find that genes whose promoters are disrupted in all three MDEMs are more likely to have dysregulated expression. Finally, we find that subtle expression changes in multiple, directly relevant genes, likely collectively contribute to the IgA deficiency and abnormal B cell development in these MDEMs.

In summary, we propose the study of MDEMs as a principled approach for systematically mapping disease-causing epigenetic variation in mammals.

PrgmNr 2911 - MeQTL Mapping for Present Cocaine Use and Persistent Cocaine Use in a Veteran Population

View session detail

Author Block: **Y. Cheng**[1], B. Li[2], A. Justice[3], K. Xu[4]; [1]Dept. of Biostatistics, Yale Sch. of Publ. Hlth., New Haven, CT, [2]Yale Sch. of Publ. Hlth., New Haven, CT, [3]Dept. of Internal Med., Yale Sch. of Med., New Haven, CT, [4]Yale Sch. Med., New Haven, CT
Disclosure Block: **Y. Cheng:** None.
Background. The influence of genetic variant on complex traits is potentially mediated through DNA methylation, referred to as Methylation quantitative trait loci (meQTLs). We recently reported significant DNA methylation aberrant in persons who persistently used cocaine. To further understand the mechanism of cocaine-associated DNA methylation, in this study, we performed a meQTL mapping for cocaine use in the Veterans Aging Cohort Study Biomarker Cohort (VACS-BC) (N=669). In this veteran population, 97.5% are male (n= 652) and 2.5% are female (n= 17), and the average age at baseline is 49.3 (SD=7.8).
Methods. Global ancestry for each sample was inferred using reference 1000 human genome. DNA methylation in whole blood samples was profiled by Illumina HumanMethylation 450K Beadchip. Present cocaine use was defined as self-reported â□□current use within 12 monthsâ□□ at a visit, and persistent cocaine use was defined as â□□current useâ□□ across 5 visits in a subset of the VACS-BC. We first conducted two separate EWAS scans for the 2 traits to select candidate CpG sites at p
Results. EWAS scans identified 497, and 499 candidate CpG sites for present cocaine use, and persistent cocaine use, respectively, while 3,058 and 4,507 meQTL clumps were further identified for the 2 corresponding traits. Nearby genes for the most significant clumped MeQTL included *MMP17* (cg22244940 - rs7487262, estimated effect = 1.59, p = 3.96e-239), *KIF25* (cg08476511 - rs58330252, estimated effect = -1.06, p = 9.42e-46), *BDNF* (cg10635145 - rs4922793, estimated effect = -0.34, p = 2.46e-17) for present cocaine use, *ATP8B4* (cg13399903 - rs7172615, estimated effect = 1.22, p = 2.37e-185) for persistent cocaine use. There were 12 and 23 pathways reaching nominal significance for the 2 traits, but they were not statistically significant after multiple test adjustments. Top ranked pathways included inflammatory bowel disease (p=0.011), cocaine addiction (p=0.014) for present cocaine use, IL-17 signaling pathway (p=0.006), viral carcinogenesis (p=0.006) for persistent cocaine use.
Conclusion. We identified a set of genetic variants significantly correlated with cocaine-use related DNA methylation. These findings deepen our understanding of genetic variants in cocaine-use related methylation alterations.

PrgmNr 2912 - Meta-analysis of gene expression data in Alzheimer's disease identifies sex-differential effects in multiple brain tissues

View session detail

Author Block: **J. Young**[1], A. Varma[1], L. Gomez[1], H. Dykxhoorn[1], B. Kunkle[1], L. Wang[2], E. R. Martin[2]; [1]Univ. of Miami, Miami, FL, [2]Univ. of Miami Miller Sch. of Med., Miami, FL
Disclosure Block: **J. Young:** None.
There are well-known epidemiologic, clinical, and neuropathological differences between men and women with Alzheimer's disease; however, the underpinnings of these differences are not well understood. Some studies point to sex differences in gene expression and the effects on AD, providing evidence of genes or pathways with sex-differential effects. We performed a meta-analysis of multiple studies, with gene expressions data on multiple brain regions to identify genes with sex-differential effects between AD cases and controls.We obtained array-based gene expression and covariate data from four datasets (GSE44772, GSE3300, GSE5281, SYN3157225). Gene expression data were normalized after stratifying datasets by tissue and sex using the Robust Multichip Averaging method. Outlier samples were detected and excluded using Principal Component Analysis. Probes with expression in the bottom 10^{th} percentile in over 80% of samples were excluded. Surrogate Variable Analysis was used to detect and adjust hidden batch effects within each dataset. Our initial analyses focused on three brain regions: dorsolateral prefrontal cortex (DLPFC 293 F/400 M), visual cortex (VC 96 F/162 M), and cerebellum (CBM 248 F/353 M). A fixed-effects inverse variance weighted meta-analysis was then used to combine cohort-specific association between AD and gene expression stratified by sex. This analysis resulted in the identification of sex-specific genes (FDR-significant exclusively in one sex). In addition, we applied a sex-interaction model that allowed us to identify genes with nominal significance in both sexes with a significant sex-interaction term. We analyzed the cross-tissue overlap of differentially expressed genes from both categories, sex-stratified and sex-interacting. We found three female-specific and 111 male-specific differentially expressed genes common to the three regions. From the sex-interacting genes, we found that two were differentially expressed in the three brain regions (*RIC8B* and *DOT1L*). Both genes showed highly significant evidence of association with AD in men and women, but the estimated effects (beta) in females in the three regions (RIC8B ranging 0.094-0.099; DOT1L ranging 0.110-0.162) were significantly larger than in males (RIC8B 0.047-0.057; DOT1L 0.026-0.047). In summary, we have performed a meta-analysis of multiple array-based gene expression studies and prioritized genes likely to have sex-differential effects on AD in multiple brain regions. Further analyses will include additional brain regions and RNAseq datasets, aiming to determine the mechanisms that underlie the difference between male and female AD patients.

PrgmNr 2913 - Population-level variation of enhancer expression identifies novel disease mechanisms in the human brain

View session detail

Author Block: **P. Dong**, G. E. Hoffman, P. Apontes, J. Bendl, S. Rahman, M. B. Fernando, B. Zeng, J. M. Vicari, W. Zhang, K. Girdhar, R. Misir, the CommonMind Consortium, K. J. Brennand, V. Haroutunian, G. Voloudakis, J. F. Fullard, P. Roussos; Icahn Sch. of Med. at Mount Sinai, New York, NY
Disclosure Block: **P. Dong:** None.
Identification of risk variants for neuropsychiatric diseases within enhancers underscores the importance of understanding the population-level variation of enhancers in the human brain. Besides regulating tissue- and cell-type-specific transcription of target genes, enhancers themselves can be transcribed. We expanded the catalog of known human brain transcribed enhancers by an order of magnitude by generating and jointly analyzing large-scale cell-type-specific transcriptome and regulome data. Examination of the transcriptome in 1,382 brain samples in two independent cohorts identified robust expression of transcribed enhancers. We explored gene-enhancer coordination and found that enhancer-linked genes are strongly implicated in neuropsychiatric disease. We identified significant expression quantitative trait loci (eQTL) for 25,958 enhancers which mediate 6.8% of schizophrenia heritability, mostly independent from standard gene eQTL. Inclusion of enhancer eQTL in transcriptome-wide association studies enhanced functional interpretation of disease loci. Overall, our study characterizes the enhancer-gene regulome and genetic mechanisms in the human cortex in both healthy and disease states.

PrgmNr 2914 - Power-improved meta-QTL analysis reveals the complex regulation of molecular QTL associated with brain diseases

View session detail

Author Block: **B. Zeng**, J. Bendl, G. E. Hoffman, R. Roussos; Icahn Sch. of Med. at Mount Sinai, New York, NY
Disclosure Block: **B. Zeng:** None.
Co-localization of risk variants for brain diseases with molecular quantitative trait loci (QTL) aims to refine the credible sets of causal variants and link them to specific genes. Increasing sample size and performing trans-ethnic analysis can increase fine-mapping resolution, and enlarge the power of colocalization analysis. We developed a statistical method, called multivariate multiple QTL (mmQTL), to perform multi-tissue and trans-ethnic large-scale QTL analysis. We first perform dense simulations to demonstrate that mmQTL increases power to detect QTLs, controls the false positive rate in trans-ethnic analysis, and reduces the credible sets of causal variants. We then applied mmQTL on brain bulk RNA-seq (n=3,956 libraries from 2,119 unique donors) and cell-type specific ATAC-seq samples (n=1995 libraries from 646 unique donors) from neuronal and glial cell types. We identify 10,769 eGenes (genes with at least a genome-wide significant eQTL), of which 5,336 eGenes have multiple eQTLs. We detect 17,425 significant caQTL in the neuronal cell, and 15,746 in the glial cell. Functional analysis based on evolutionary scores and rare-variant burden reveals eGenes are more likely to be mutant tolerant than no-eQTL genes, indicating evolutionary constraints on gene expression in the brain. Non-primary eQTLs capture regulatory elements defined by ATAC-seq peaks and are enriched for neuro-degenerative diseases. We found that caQTLs are largely shared among brain regions, but not in cell types. Fine-mapped variants from the power-improved caQTL are enriched to disrupt TF binding sites, and there are different sets of TF genes affected in neuronal and glial cells. Lastly, we integrated the detected eQTL and caQTL to conduct caQTL-eQTL-GWAS colocalization analysis, and revealed the regulation patterns for some disease-associated genes in the human brain. In this study, we have created the largest brain eQTL and caQTL resources to date, and they should be valuable to the community in exploring the underlying biological mechanism in brain diseases.

PrgmNr 2915 - Sex differences in the human brain transcriptome of cases with schizophrenia

View session detail

Author Block: **Y. Ma**; Icahn Sch. of Med. at Mount Sinai, New York City, NY
Disclosure Block: **Y. Ma:** None.
Background: Schizophrenia differs between males and females in the course of the disease and the molecular mechanisms underlying these differences remain uncharacterized. Schizophrenia is a multi-factorial neurodevelopmental impairment of the brain that can be attributed to both genetic and environmental factors. Gene expression is a consequence of both the genetic and environmental factors that contribute to the pathophysiology of the disease. **Methods:** In order to address questions about how sex differences contribute to schizophrenia gene expression, we performed a large-scale multiscale embedded-gene coexpression network analysis (MEGENA) of RNA-seq data from 437 controls and 341 cases from two distinct cohorts from the CommonMind Consortium. **Results:** By performing a network analysis to reduce dimensionality and elucidate interactions among genes, we identified co-expressed gene modules and then characterized their enrichment for differential expression signatures for diagnosis, sex, and sex-by-diagnosis interaction. Sex-by-diagnosis interaction associated modules were more likely to be enriched for diagnosis differential expression signatures for both cohorts. Only a small number of modules were enriched for sex signatures, which show only partial overlap with the diagnosis effects. We found enrichment of co-expression modules for sex-by-diagnosis differential expression signatures, which were highly reproducible across the two cohorts and involve a number of diverse pathways, including neural nucleus development, neuron projection morphogenesis, and regulation of neural precursor cell proliferation. For the top sex-by-diagnosis interaction enriched module, we identified a significant enrichment for schizophrenia common genetic variation. CASKIN1 is a key regulator of this module that is reproduced among both cohorts. CASKIN1 has a reduced expression in schizophrenia patients and is also known as a synaptic scaffolding protein to play a role in signal transduction. **Conclusions:** The network approach has uncovered a number of gene modules and key regulators of schizophrenia. Our results indicate that any sex differences in schizophrenia gene expression signatures are likely small. This underscores the challenge of identifying robust sex-by-diagnosis signatures, which will require additional analyses in larger cohorts in the future.

PrgmNr 2916 - Sex-specific profiles of m6A RNA methylation in the brain of individuals with major depressive disorder

View session detail

Author Block: **H. Mitsuhashi**[1], C. Nagy[1], Z. Aouabed[2], P. Danthi[2], G. Turecki[3]; [1]McGill Univ., Montreal, QC, Canada, [2]McGill Group for Suicide Studies, Montreal, QC, Canada, [3]Mc Gill Univ, Montreal, QC, Canada
Disclosure Block: **H. Mitsuhashi:** None.
Introduction: Females are twice as likely to be diagnosed with Major Depressive Disorder (MDD); however, males are 3.5 times more likely to die by suicide. This is a striking example of sex differences in MDD, and mounting evidence suggests that it may be driven by sex-specific molecular mechanisms. Epigenetic mechanisms, which are altered in response to environmental factors, are known to be involved in the pathophysiology of MDD; however, little is known about the impact of the epitranscriptome. In recent years, RNA modifications have emerged as a dynamic and crucial mechanism in the post-transcriptional regulation of gene expression. Among the 150 known RNA modifications, N6-methyladenosine (m6A) is the most abundant and reversible RNA modification in mammalian messenger RNA (mRNA). Emerging evidence suggests that m6A plays an important role in the brain, including neurodifferentiation, neurogenesis, and memory and learning. Moreover, recent studies have linked m6A to molecular and behavioral responses to stress, making it an important candidate regulator of stress-related psychiatric disorders, including MDD. This study aims to describe the landscape of m6A in the human brain and to identify changes that may occur in the context of MDD. **Methods:** First, the postmortem stability of m^6A and the influence of age, RNA Integrity, and pH on global m^6A levels were investigated in human postmortem brain tissue. Next, we optimized a low-input m6a-seq method for human postmortem brain tissue and confirmed that the m6a-peaks are enriched in the known m6A GRACH motif and near 3â□□UTR and stop codon as suggested by the previous study. The ventromedial prefrontal cortex was obtained from male and female MDD and healthy control subjects to investigate the role of m6A in MDD. We performed m6A-seq and RNA-sq to investigate m6A at transcripts levels and the impact of m6A on gene expression. **Results:** Our results suggest that PMI does not significantly influence global m^6A levels, and m^6A is relatively stable in the postmortem brain. In our m6a-seq data, we identified ~25,000 m6A peaks in the human brain, and these peaks were enriched in genes related to neuronal and synaptic regulation. Moreover, our results show a distinct m6A profile in MDD and control, with a little overlap between males and females. **Conclusion:** This project will help us understand the role of m6A in stress-related psychiatric disorders and will serve as a much-needed example of sex-specific analysis in psychiatric research.

PrgmNr 2917 - Single Nuclei Sequencing of Human Putamen Oligodendrocytes Reveals Altered Heterogeneity and Disease-Associated Changes in Parkinsons Disease and Multiple System Atrophy

View session detail

Author Block: **E. Teeple**[1], P. Joshi[2], R. S. Pande[1], Y. Huang[1], A. Karambe[1], M. Latta-Mahieu[2], P. Sardi[3], A. Cedazo-Minguez[2], K. W. Klinger[4], A. Flores-Morales[2], S. L. Madden[1], D. Rajpal[1], D. Kumar[5]; [1]Sanofi, Framingham, MA, [2]Sanofi, Chilly-Mazarin, France, [3]Genzyme a Sanofi Company, Framingham, MA, [4]Sanofi Genzyme, Framingham, MA, [5]Framingham, MA

Disclosure Block: **E. Teeple:** Salary/Employment; Sanofi.

The role of oligodendrocytes in neurodegenerative diseases remains incompletely understood. We profiled 87,086 single nuclei from human brain putamen region for healthy control, Parkinsons Disease (PD), and Multiple System Atrophy (MSA). Oligodendrocyte lineage cells were the dominant celltype in the putamen and subpopulations clustered by transcriptomic variation found to exhibit diverse functional enrichment patterns. Oligodendrocyte heterogeneity was altered in disease-specific ways. Among profiled subpopulations, differences in expression of SNCA, HAPLN2, MAPT, APP, and OPALIN were observed for PD and MSA compared with healthy controls. Intriguingly, greater activation of unfolded protein response pathway gene expression was observed in PD versus MSA. Using network analysis, we identified specific PD- and MSA-correlated gene co-expression modules enriched with disease-relevant pathways, and the PD-correlated module was significantly enriched for Parkinsons Disease GWAS loci ($p = 0.01046$). Our analysis provides a broader understanding of oligodendrocyte functional biology and reveals distinctive oligodendrocyte pathological alterations associated with PD and MSA which may suggest potential novel therapeutic targets and new strategies for disease modification.

PrgmNr 2918 - TET1 mediated modulation of Alzheimerâ□□s disease

View session detail

Author Block: **M. Armstrong**, P. Jin; Emory Univ., Atlanta, GA
Disclosure Block: **M. Armstrong:** None.
Background: Human sequencing studies have identified roughly 30 risk loci associated Alzheimerâ□□s disease (AD). However, these loci only explain a portion of disease pathogenicity. Recent studies support a link between gene-environment interactions and AD, and published work from our lab indicates *TET1* mediates gene-environment interactions and *TET1* KO reduces stress response in mice [1]. Furthermore, DNA deep sequencing studies identified an enrichment of rare *TET2* variants associated with AD [2]. This study examines how knockout of the *TET1*, a DNA 5-hydroxymethylcytosine (5hmC) regulator, influences AD-associated transcription and pathological outcomes in the 5xFAD AD mouse model.
Methods: We measured cognitive/behavioral, 5hmC, and expression profiles of a *TET1* heterozygous knockout in the 5xFAD mouse model to examine the influences of *TET1* KO on AD pathogenicity.
Results: We show significant differences in behavior, cognition, 5hmC methylation and expression profiles of FAD/$TET1^{+/-}$ mice relative to WT and FAD. In relation to FAD, FAD/$TET1^{+/-}$ improved stress response in the Tail-Suspension Assay ($p = 0.0438$), and relative to WT mice, FAD/$TET1^{+/-}$ mice display a reduction in their latency to mount the platform in the Morris water maze ($p = 0.01$). Analysis of 5hmC capture data revealed 2536 differentially hydroxymethylated regions (DhMRs) upregulated in FAD/*TET1*+/- and 410 DhMRs downregulated relative FAD (FDR > 0.05). Analysis of the expression data highlighted 132 differentially expressed genes (DEGs) in FAD/$TET1^{+/-}$ relative to FAD (FDR > 0.05). Among the genes exclusive to the FAD/*TET1*+/- mice were *NR4A1*, *ARC*, and *APLP1*. Notably, *NR4A1* appeared in both the 5hmC and RNA-Seq data with increased levels of 5hmC and gene expression relative to FAD mice. Gene ontology analysis of RNA expression data identified significant changes in pathways involved in regulation of ion transmembrane transport, behavioral fear response, regulation of trans-synaptic signaling, and learning or memory.
Conclusion: Taken together, these results suggest *TET1* modulates AD-associated gene regulation, expression, and pathology.
References: 1. Cheng Y, et al., â□□Ten-Eleven Translocation Proteins Modulate the Response to Environmental Stress in Mice.â□□ *Cell Reports*, 2018. 2. Cochran JN, et al., Non-coding and Loss-of-Function Coding Variants in *TET2* are Associated with Multiple Neurodegenerative Diseases. *American Journal of Human Genetics*, 2020.

PrgmNr 2919 - The Musculoskeletal 3D Epigenome Atlas

View session detail

Author Block: **M-J. Tsai**[1], S. Reppe[2], T. Sato[3], R. Gill[4], M. Wein[3], K. Gautvik[2], NHLBI TOPMed Bone Mineralization workgroup, Y-H. Hsu[1]; [1]Marcus Inst. for Aging research and Harvard Med. Sch., Boston, MA, [2]Oslo Univ. Hosp., Oslo, Norway, [3]Massachusetts Gen. Hosp. and Harvard Med. Sch., Boston, MA, [4]Genentech, San Mateo, CA

Disclosure Block: **M. Tsai:** None.

Musculoskeletal (MSK) disorders are a common and costly problem for elderly populations worldwide, most of which are polygenic disorders. Although an increasing number of human genome-wide association studies (GWAS) and population-based biobank studies identified sequence variants associated with risks of musculoskeletal diseases, there is a lack of information about the cell types in human bone and skeletal muscle tissues. Moreover, it is not a trivial task to understand specific genes regulating the landscape and understanding their functional involvement in musculoskeletal biology, especially for non-coding related variants.To address this issue, we generated HI-C seq, ATAC-seq, and RNA-seq in human primary mesenchymal stem cells, pre-mature osteoblasts, matured osteoblasts, osteocytes, skeletal myoblasts, and myotubes obtained from bone and skeletal muscle biopsies. These data were then combined with publicly available DNase-seq and ChIP-seq in a handful of relevant cell types from the ENCODE and Roadmap Epigenomics projects. We established global maps of regulatory elements and 3D chromatin looping structure with high-resolution (2kb) in human MSK relevant primary cell types. We identified (1) MSK-specific active enhancer-like regions; (2) MSK-specific active transcription factor binding sites and downstream-regulated genes in the open chromatin regions; (3) MSK-specific active promoter-like regions and transcription factors for protein-coding genes; (4) MSK specific proximal and distal enhancer-promoter interactions via the high-resolution 3D loop structure with ATAC-seq and RNA-seq data. The 3D MSK gene regulatory circuits and landscapes are implemented into an online searchable browser, which provides target gene prediction for all available (~700 million) non-coding variants for relevant diseases, visualization for MSK cell-type-specific gene regulatory circuits annotation for every sequence variant observed so far. We also integrated all update-to-date GWAS findings from GWAS Catalog to provide post-GWAS analyses for in-silico functional annotation of associated variants and underlying causal variants and genes prediction. We plan to extend our 3D Epigenome Atlas with additional human primary musculoskeletal cell types with single-cell RNA-seq and single-cell ATAC-seq.In summary, by integrating cell-type-specific multi-omics data, we have established MSK gene regulatory landscapes and developed an online searchable browser that provides comprehensive epigenome annotation and visualization of the 3D genome chromatin interactions. The browser is publicly available.

PrgmNr 2920 - Unraveling the role of cell type specific noncoding variations in craniofacial disease

View session detail

Author Block: **E. Winchester**[1], T. Yankee[2], K. Child[3], J. Cotney[3]; [1]UConn Hlth.DMD/PhD Program, Farmington, CT, [2]UConn Hlth.Graduate Program, Farmington, CT, [3]UConn Hlth.Dept. of Genetics and Genome Sci., Farmington, CT
Disclosure Block: **E. Winchester:** None.
Craniofacial disorders such as orofacial clefting are common congenital defects which affect a significant percent of live births, posing a public health and financial burden to society. Previously, we have associated the role of noncoding genomic sequences of the developing face in the heritability of craniofacial morphology and oral clefting. The noncoding genome contains sequences called enhancers which facilitate expression of target genes through recruitment and binding of transcription factors. Enhancers are typically active in developmental stage and cell type specific patterns, controlling the spatiotemporal expression patterns of target genes. Enhancer sequence variation can alter expression of target genes to cause isolated phenotypes in the cell type of activity, ranging from normal morphological differences to malformations. While we have established the role of craniofacial-specific enhancers in normal morphology and disease through bulk craniofacial assays, the contributing cell types are unknown, limiting work in prevention and treatment of malformations. To uncover cell type-specific enhancers whose sequence variants contribute to normal facial morphology and malformations, we profiled regions of the genome that were accessible at the single nucleus level (snATAC-seq) in primary human craniofacial tissues from 7 distinct Carnegie stages (CS12-20; 4-9 weeks post conception). In our initial analysis of CS17, we identified 6 distinct cell types and their unique regions of accessible chromatin. Single nucleotide polymorphisms associated with orofacial clefting and normal facial morphological differences were interrogated for their enrichment in these cell-type specific accessible regions of chromatin and contrasted to similarly profiled regions from hundreds of tissues from the Encode Consortium. We observed that variants associated with normal craniofacial morphology are significantly enriched in mesenchyme-specific accessible regions. Interestingly, variants that are associated with risk for orofacial clefting showed differences in cell-type specificity based on the population ancestry; in studies of individuals of European ancestry, clefting variants were enriched in mesenchyme-specific regions while in East Asian ancestry variants were enriched in epithelial regions. These results indicate differences in molecular mechanisms underlying similar phenotypes in different populations. In further analysis of the remaining stages we will confirm these results and refine the window of human craniofacial development and specific cell types that contribute to craniofacial development and malformation.

PrgmNr 2921 - Widespread shifts in muscle gene expression associated with reduced gait speed in a non-human primate model of aging

View session detail

Author Block: **E. E. Quillen**, B. Frye, G. Li, J. Chan, M. Stainback, T. Register, S. Craft, L. Cox, C. Shively; Wake Forest Sch. of Med., Winston Salem, NC
Disclosure Block: **E.E. Quillen:** None.
Declining muscle mass leading to sarcopenia is a major contributor to increased morbidity and loss of independence among older adults. Gait speed is a reliable biomarker of overall musculoskeletal health as well as increased risk of disability and mortality among both humans and non-human primates. In a cohort of 28 female, middle-aged and older adult vervet monkeys (*C. sabaeus*) housed in the Vervet Research Colony at Wake Forest School of Medicine, we measured habitual gait speed and generated total RNA sequence data from contemporaneous muscle biopsies of the vastus lateralis. We focus here on mRNA. Following quality control, reads were aligned to the current version of the vervet transcriptome and lowly expressed genes filtered out resulting in > 9,000 transcripts which were TMM normalized and evaluated for association with gait speed while controlling for age. To account for the genetic relatedness of the vervets which are part of a larger pedigree, we fit a generalized linear mixed model including kinship as a random effect. 368 transcripts were significantly differentially expressed, of which 67% were upregulated in association with faster gait speed. The genes showing the greatest log-fold changes were also those with the highest average expression including genes encoding major contractile proteins in fast twitch (*MYH1*) and slow twitch muscle fibers (*TNNT1*, two transcripts). Additionally, major shifts were identified in the expression levels of energy metabolism genes including enzymes that catalyze ATP hydrolysis (*ATP2A3*), iron metabolism *(FTH1*), and mitochondrial activity (*IDH2*). To contextualize the 368 significantly associated genes, we compared them to previously generated muscle transcriptome data sets collated from the Gene Expression Omnibus (GEO) by SysMyo. We performed Fisherâ□□s exact tests to compare the enrichment of our 368 differentially expressed genes to 1000 randomly generated sets of 368 genes present in muscle. Empirical *p*-values

PrgmNr 2922 - Accurate identification of circRNA landscape and dynamic regulation during early neuronal differentiation

View session detail

Author Block: **F. Wang**, Y. Li, B. Yao; Dept. of Human Genetics, Emory Univ. Sch. of Med., Atlanta, GA
Disclosure Block: **F. Wang:** None.
Circular RNAs (circRNAs) are a class of endogenous single-stranded RNAs with a unique circularized structure derived mostly from back-splicing of exons in precursor mRNAs and have recently emerged as one of the pivotal regulatory RNAs in mammals. CircRNAs has been shown to undergo spatiotemporal regulation and play important roles during mammalian brain development. However, circRNA identification and quantification have been suffering from relatively low power and sensitivity, and prone to high false discovery rate at both experimental and computational level. Here, we employed a recent-established method in combination of poly(A) tailing and RNase R digestion (refer as â□□A-tailingâ□□) to effectively deplete linear RNAs for circRNA enrichment. By using A-tailing approach coupled with rRNA-depleted RNA-seq, we interrogated the genome-wide circRNAs landscape and their dynamic regulation during early neuronal cell differentiation using human neuroblastoma cell line SH-SY5Y. SH-SY5Y cells were assayed in a 17-day differentiation protocol by retinoic acid and BDNF treatment. Successful differentiation was validated at both cellular and molecular levels, with the differentiated cells showing expansive and branched neuronal phenotype and expression of neuronal markers including TUJ1, MAP2 and SMI31. Comparing with published circRNA profile in SH-SY5Y cells, our data displayed robust, sensitive and high confident identification that captured more than 19,000 previously undetected circRNAs in SH-SY5Y cells. During SH-SY5Y differentiation, 404 and 619 circRNAs were significantly downregulated or upregulated, respectively (FDR

PrgmNr 2923 - Alcohol Use Disorder is associated with DNA methylation-based shortening of telomere length and regulated by *TESPA1*: implications for aging

View session detail

Author Block: **J. Jung**, J. Wagner, F. W. Lohoff; Natl. Inst. on Alcohol Abuse and Alcoholism, Bethesda, MD
Disclosure Block: **J. Jung:** None.
Chronic heavy alcohol consumption is associated with increased mortality and morbidity and often leads to premature aging; however, the mechanisms of alcohol-associated cellular aging are not well understood. In this study, we used DNA methylation derived telomere length (DNAmTL) as a novel approach to investigate the role of alcohol on the aging process. DNAmTL was estimated by 140 cytosine phosphate guanines (CpG) sites in 372 individuals with alcohol use disorder (AUD) and 243 healthy controls (HC), including various endophenotypes and clinical biomarkers. Exploratory genome-wide association studies (GWAS) on DNAmTL were performed to identify genetic variants contributing to DNAmTL shortening. Top GWAS findings were analyzed using in-silico expression quantitative trait loci analyses and structural MRI hippocampus volumes of individuals with AUD. DNAmTL was 0.11-kilobases shorter per year in AUD compared to HC after adjustment for age, gender, race, blood cell counts (p=4.0x10^{-12}). This association remained significant after additionally adjusting for BMI, and smoking status (0.05 kilobases shorter per year, p=0.002). DNAmTL shortening was strikingly associated with chronic heavy alcohol use (psTESPA1), at the genome-wide level (p=3.87x10^{-8}). The major allele C of rs4374022 driving the DNAmTL shortening was associated with decreased hippocampus volume (pp=0.04). Our study demonstrates DNAmTL-related aging acceleration in AUD and suggests a functional role for *TESPA1* in regulating DNAmTL possibly via the immune system with subsequent biological effects on key brain regions implicated in aging.

PrgmNr 2924 - Chromatin accessibility and gene expression during adipocyte differentiation identify context-dependent effects at cardiometabolic GWAS loci

View session detail

Author Block: **H. J. Perrin**[1], K. W. Currin[2], S. Vadlamudi[1], G. Pandey[3], K. K. Ng[1], M. Wabitsch[4], M. Laakso[5], M. Love[6], K. L. Mohlke[2]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Univ North Carolina, Chapel Hill, NC, [3]Carrboro, NC, [4]Univ. of Ulm, Ulm, Germany, [5]Univ. of Eastern Finland and Kuopio Univ. Hosp., Kuopio, Finland, [6]UNC-Chapel Hill, Chapel Hill, NC
Disclosure Block: **H.J. Perrin:** None.
Chromatin accessibility and gene expression in relevant cell contexts can guide identification of regulatory elements and mechanisms at genome-wide association study (GWAS) loci. To identify regulatory elements that display differential activity across adipocyte differentiation, we performed ATAC-seq and RNA-seq in a cell model of preadipocytes and of adipocytes at days 4 and 14 of differentiation. For comparison, we created a consensus map of ATAC-seq peaks in 11 subcutaneous adipose tissue samples. We identified 58,387 context-dependent chromatin accessibility peaks and 3,090 context-dependent genes between all timepoint comparisons (log2 fold change>1, FDR

PrgmNr 2925 - Deconvolution of genetic variation using high-quality *cis*-regulatory elements map of kidney cells

View session detail

Author Block: **S. Han**[1], Y. Muto[2], P. C. Wilson[2], B. D. Humphreys[2], M. G. Sampson[1], D. Lee[1]; [1]Boston Children's Hosp., Boston, MA, [2]Washington Univ. in St. Louis, St. Louis, MO
Disclosure Block: **S. Han:** None.
Genome-wide association studies (GWAS) have facilitated the discovery of disease- or trait-associated genetic variants that can ultimately lead to improved precision of clinical diagnosis and/or molecular pathogenesis in a translational medicine framework. However, identifying specific cell types within organs in which the GWAS variants exert their function remains a significant challenge, especially for the complex and heterogeneous kidney. To tackle this, we constructed high-quality maps of cis-regulatory elements (CREs) for kidney cells to deconvolute GWAS variants for kidney-relevant phenotypes. Specifically, we devised a computational framework using a sequence-based predictive model that maximally detects CREs by identifying open-chromatin regions with marginal read-mappings but harboring CRE sequence features. We applied this method to kidney ATAC-seq data. Our high-quality CRE maps have enabled us to detect >100,000 CREs for podocytes, a key rare (h_g^2]=9.3%). Heritability analysis using these CRE maps uncovered the differential contribution of specific cell types to two major kidney functional traits, UACR and estimated glomerular filtration rate (eGFR). As would be predicted from physiologic understanding, CREs for podocytes and proximal tubule cells (PT) had enriched proportion of SNP-heritability for UACR and eGFR, respectively (UACR: Pr[h_g^2]/Pr[SNPs]=6.8 for podocyte, 2.3 for PT; eGFR: Pr[h_g^2]/Pr[SNPs]=-1.9 for podocyte, 4.3 for PT. Moreover, we found the podocyte relevance of a known GWAS variant (rs17831251; OR=2.25, P=4.7Ã□10$^{â□□103}$) on *PLA2R1* associated with Membranous Nephropathy. Our CRE map showed strong podocyte-unique CRE that overlaps with the index variant, suggesting that the index SNP is potentially the causal variant perturbing podocyte-specific transcriptional regulation of *PLA2R1*. Taken together, we expect that the deconvolution of GWAS variants using the high-quality kidney CRE maps will provide cell-type relevance of GWAS variants on genetic effects not captured by single-cell RNA-seq alone. This will empower the functional interpretation of genetic variation associated with kidney traits and diseases.

PrgmNr 2926 - Engineering a synthetic humanized *RET* mouse to model Hirschsprung Disease

View session detail

Author Block: **R. Fine**[1], S. Chatterjee[1], J. Laurent[2], J. Boeke[2], A. Chakravarti[1]; [1]NYU Sch. of Med., Ctr. for Human Genetics and Genomics, New York City, NY, [2]NYU Sch. of Med., Inst. for Systems Genetics, New York City, NY

Disclosure Block: **R. Fine:** None.

Many human disorders are genetically multifactorial with no necessary and sufficient relationship between genotype and phenotype, as compared to Mendelian traits. An exception is Hirschsprung Disease (HSCR) which is characterized by the failure of enteric neural crest cells to colonize and innervate the intestinal tract, leading to bowel obstruction and death if left surgically unrepaired. Previous sequencing studies by our group have revealed that the majority of HSCR risk arises from coding and transcriptional regulatory variants in the gene encoding the tyrosine kinase receptor, RET, and its transcriptional gene regulatory network (GRN). Here, hypomorphic regulatory (enhancer) variants have lower penetrance, but higher risk, than pathogenic coding variants. Additionally, coding sequence variants in another 33 genes and loci with *RET* contribute to ~67% of population attributable disease risk. To understand the disproportionate contribution of *RET* coding and enhancer variants to HSCR, we are building mouse models with an intact *RET* human genomic locus. This enables direct modeling of coding and regulatory human variants in mice without the need for phylogenetic conservation. Our approach is to use â□□Big-DNAâ□□ technology in budding yeast and *E. coli* to synthetically construct these loci in a parallel manner. To date, we have constructed several loci of 60, 80, and 179kb lengths with defined haplotypes for the entire *RET* gene body and three major transcriptional enhancer HSCR-associated SNPs denoted as rs2506030, rs7069590, and rs2435357. We then use highly engineered Inducible Cassette Exchange â□□ICEâ□□ targeting vectors to insert these synthetic loci into mouse embryonic stem cells and ultimately derive mice. We are using the mouse genetic toolbox to cross-breed additional hypomorphic alleles of a second major node in the *RET* GRN, *EDNRB*, to test epistatic interactions. We have also established 5C-ID sequencing in CHP212 cells and discovered topological chromatin interactions between the three aforementioned enhancer SNPs and the RET promoter. We will combine this approach with cutting-edge single cell technologies in mice to understand the genetic rules by which variants specify regulatory changes to alter gene expression and, in turn, affect cell behavior and clinical penetrance of disease.

PrgmNr 2928 - Functional characterization of regulatory elements involved in mouse CD4$^+$ T cell differentiation

View session detail

Author Block: **K. Siklenka**, L. Li, A. Barrera, M. Parker, R. Venukuttan, Y-S. Kim, M. Gemberling, S. Snyder, C. A. Williams, G. E. Crawford, C. Gersbach, M. Ciofani, T. E. Reddy; Duke Univ., Durham, NC
Disclosure Block: **K. Siklenka:** None.
The adaptive immune system is an exceptional model for studying complex cellular phenotypes with a high degree of therapeutic significance. CD4$^+$ T helper cells are important mediators of both immunity and autoimmunity. With unique functions, T helper cells support other cell types in clearing of infection or by attenuating an immune response. Their dysregulation is often associated with chronic inflammation, autoimmune disease, and allergy. Over the past two decades, major advancements in genetics and genomic technologies have revealed that combinations of extrinsic and intrinsic factors govern T helper cells activation, differentiation, and immune function. Now, many datasets have detailed the genomic location of thousands of candidate regulatory elements involved in these processes; however, the functional contribution of those elements remain poorly understood and difficult to study. **Objective:** To empirically measure the activity and function of thousands of candidate regulatory elements involved in the differentiation of mouse T helper subsets. This comprehensive functional study of regulatory elements in primary immune cells will enable the characterization of both the genetic and epigenetic components responsible for regulating gene expression during an immune response. We anticipate these results will lead to improved design of therapeutic interventions for immune-related disease. **Approach:** We used the high-throughput reporter assay STARR-seq coupled with ATAC-seq to assay the regulatory activity of all open chromatin regions in mouse T helper type 1 (Th1), type 2 (Th2), type 17 (Th17), and T regulatory (Treg) cells *in vitro*. We applied a joint analysis model to those multiple STARR-seq datasets to identify active regulatory elements that are common or unique across each subtype. We analyzed sequence features and integrated gene expression and epigenetic datasets to determine what core motifs and chromatin context drive regulatory activity. We applied an orthogonal CRISPR-interference screen in mouse Th17 cells by targeting dCas9-KRAB to a subset of candidate regulatory elements during differentiation. We used FACS to enrich for cells containing gRNA that cause altered expression of the Th17 defining genes *Rorc* and *Il17a*. Using a transgenic T cell transfer model of Th17 differentiation coupled with dCas9-KRAB we targeted individual elements identified *in vitro* and confirmed their essentiality *in vivo* by characterizing cell phenotypes with flow cytometry. Together these results identify a set of functional regulatory elements that may offer an expanded understanding of the active regulome that governs an adaptive immune response.

PrgmNr 2929 - Human microglia regulome analysis defines the inherited risk loci in Alzheimerâ□□s Disease

View session detail

Author Block: **R. Kosoy**[1], B. Zeng[1], B. Jaroslav[1], P. Dong[1], S. Rahman[1], S. Kleopoulos[1], T. Raj[1], J. Humphrey[1], K. de Paiva Lopes[1], C. Alexander[1], C. Kellner[1], V. Haroutunian[1,2], G. E. Hoffman[1], J. Fullard[1], P. Roussos[1,2]; [1]Icahn Sch. of Med. at Mount Sinai, New York City, NY, [2]JJ Peters VA Med. Ctr., New York City, NY

Disclosure Block: **R. Kosoy:** None.

Microglia are the brain resident macrophages with a range of both immune and neuronal maintenance related functions. They have been implicated as the main cell type contributing to Alzheimerâ□□s Disease etiology. The transcriptional regulation in human microglia has not been well explored thus far, limiting our understanding of the molecular and cellular mechanism in the AD etiology. Here we present a comprehensive and unique transcriptome and chromosome accessibility landscape atlas in primary human microglia (146 samples) representing both healthy and neurodegenerative states. Transcription factor (TF) regulatory network analyses identified SPI1, IRF1, and PURA as the most relevant TFs to AD. Utilizing the in-house generated microglia HiC data, we applied the Activity-By-Contact (ABC) method and identified 24,459 enhancer-promoter interactions which are highly enriched for AD common variants. We then explored the genetic regulation of transcription and chromatin accessibility and identified 7,302 eQTLs and 10,266 chromosome accessibility QTLs (caQTLs) (q

PrgmNr 2930 - Identification of bidirectional regulatory region between *FADS1* and*FADS2*

View session detail

Author Block: **S. Yang**[1], K. Ye[2]; [1]Dept. of Genetics, Univ. of Georgia, Athens, GA, [2]Univ. of Georgia, Athens, GA
Disclosure Block: **S. Yang:** None.
The *FADS1* and *FADS2* genes encode two fatty acid desaturases which act as rate-limiting enzymes in the biosynthesis process of long-chain polyunsaturated fatty acids (LC-PUFAs). These two genes are located on chromosome 11 in a head-to-head orientation, resulting in a potential shared regulatory region between them. Based on previous literature, this region could be a potential â□□bidirectional regulatory regionâ□□, which may regulate the gene expression on opposite direction simultaneously. Many studies have revealed that there were certain features harbored by such bidirectionally functional region, many of which are shown in the shared region of *FADS1* and *FADS2*, including the existence of CpG islands, lack of a typical TATA box, being a pair of genes arose by gene duplication events, and shared transcription factor binding sites. Furthermore, recent analysis using GTEx (Genotype-Tissue Expression) data also revealed that some identified eQTLs were uniformly found to be associated in contrary directions for *FADS1* and *FADS2* expression, which suggests that genetic variants within the *FADS* cluster may affect the transcription of these two genes in an opposite way. This inverse expression may be explained by steric hindrance caused by the binding of transcription factors targeting expression of one of the two genes. All these existing lines of evidence indicate the bidirectional regulation is possible in that shared region. Therefore, the purpose of this study is to investigate whether and how the shared region affect expression of the two genes. To our knowledge, this is the first study on the bidirectional regulatory region between *FADS* genes. If successfully elucidated, findings from this project will enhance not only our understanding of underlying mechanism of *FADS* expressions, but also bidirectional regulation of genes. In the long term, this project will also facilitate the development of genome-informed LC-PUFAs supplementation for preventing associated complex diseases.

PrgmNr 2931 - Incorporating local ancestry improves the identification of ancestry-associated methylation signatures and meQTLs in the African American population

View session detail

Author Block: **B. Li**[1,2], A. C. Justice[3,2], B. E. Aouizerat[4], K. Xu[5], H. Zhao[6]; [1]Yale Sch. of Publ. Hlth., New Haven, CT, [2]VA Connecticut Hlth.care System, West Haven, CT, [3]Yale Sch. of Med., New Haven, CT, [4]Bluestone Ctr. for Clinical Res., New York Univ., New York, NY, [5]Yale Sch. Med., New Haven, CT, [6]Yale Univ. Sch. of Publ. Hlth., New Haven, CT
Disclosure Block: **B. Li:** None.
It is well known that there are widespread methylation differences across self-reported racial/ethnic groups in the human genome. However, few studies have aimed to identify ancestry-associated methylation marks in admixed populations. For admixed samples, the demographic information collected from participants is inadequate to precisely define population structure. Genetically inferred admixture provides a more accurate proxy of the heterogeneity of genetic admixture. We carried out three epigenome-wide association studies (EWASs) of DNA methylation on genetic admixture (local and global ancestry) and self-identified race/ethnicity in the Veterans Aging Cohort Study (VACS) (N=994). We identified 708, 30, and 1,284 CpGs significantly associated with self-reported race/ethnicity, global ancestry, and local ancestry, respectively. In replication analyses, the proportion of replicated CpGs associated with local ancestry was greater than those identified for global ancestry and race/ethnicity. Importantly, the identified ancestry-associated CpGs show a high SNP-based heritability (mean heritability=0.41) These CpGs are also significantly depleted in promoter regions and CpG islands while moderately enriched in the south shore of CpGs in the methylome. Furthermore, by incorporating SNP-based estimates of ancestry, we identified more methylation quantitative trait loci (meQTL) clumps for ancestry-associated methylations (N=2,352) than those identified from the model that assumes identical genetic effects across ancestry origins (N=1,784). The majority (71.4%) of meQTLs displayed significantly different genetic effects in the context of an African or European ancestry background. Inferred local ancestry can better characterize ancestral contributions to genetic architecture across the genome. Incorporating local ancestry information can better identify ancestry-enriched methylation in admixed populations. Determining meQTL by incorporating local ancestry can reveal ancestry-specific genetic effects on methylation. Taken together, these findings have important implications for conducting EWAS in admixed samples in the multi-ancestry cohorts.

PrgmNr 2932 - Sex differences in the intergenerational link between maternal and neonatal whole blood DNA methylation: An analysis in the Boston Birth Cohort

View session detail

Author Block: **J. Hu**[1,2], X. Xu[3], J. Li[4], X. Hong[5], K. M. Rexrode[1], F. B. Hu[6,7], X. Wang[5,8], L. Liang[4]; [1]Div. of Women's Hlth., Dept. of Med., Brigham and Women's Hosp., Boston, MA, [2]Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [3]Dept. of Biostatistics, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [4]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [5]Ctr. on the Early Life Origins of Disease, Dept. of Population, Family and Reproductive Hlth., Johns Hopkins Univ. Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [6]Dept.s of Nutrition and Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [7]Channing Div. of Network Med., Dept. of Med., Brigham and Women's Hosp., Boston, MA, [8]Div. of Gen. Pediatrics & Adolescent Med., Dept. of Pediatrics, Johns Hopkins Univ. Sch. of Med., Baltimore, MD

Disclosure Block: **J. Hu:** None.

Intergenerational inheritance of DNA methylation (DNAm) variations could contribute to the inheritance of disease susceptibility across generations. No study has investigated patterns of intergenerational transmission of DNAm at the genome-wide scale. It is also unknown whether there are sex differences in maternal-neonatal DNAm patterns. This study included 396 mother-newborn pairs (45.5% male) from the Boston Birth Cohort. Genome-wide DNAm profiling was performed on maternal whole blood (obtained 24-72 hours after delivery) and neonatal cord blood samples using the Illumina MethylationEPIC BeadChip; 721,395 DNAm sites on autosomes and the X chromosome (ChrX) were eligible for analysis after quality control steps, including removing probes mapped to multiple genomic regions. **First**, using Spearmanâ□□s rank correlation, we estimated intergenerational concordance in genome-wide DNAm and found significant sex differences in mother-newborn correlations in DNAm patterns ($p=3.8Ã□10^{-8}$), with female newborns having relatively stronger correlations. Sex differences in correlations were attenuated but remained significant after excluding DNAm sites on ChrX (p=0.035). **Second**, we estimated sex differences in associations between maternal and neonatal methylation levels at each DNAm site using linear regressions adjusting for maternal age at delivery, maternal race/ethnicity, type of delivery, preterm delivery, maternal smoking, and surrogate variables. A significant interaction between newbornâ□□s sex and maternal DNAm was observed for 32 DNAm sites (FDREFHC2, *PCDH19*, *TSPYL2*, etc.) and the other 10 were mapped to 8 genes on autosomes (*ADCYAP1R1*, *UHRF1*, *KLF13*, etc.). In addition, 18 of these 32 DNAm sites showed stronger intergenerational correlations for female newborns. **Finally**, we calculated mother-newborn differences in methylation levels at individual DNAm sites and estimated their associations with newbornâ□□s sex using linear regressions adjusting for the same set of covariates. We observed 27,887 DNAm sites that showed significant associations with newbornâ□□s sex (FDRRBM3, *COL4A6*, *TSPAN7*, etc.), and the top 10 autosomal DNAm sites were mapped to 8 genes (*LOC644649*, *UBTF*, *FOXN3*, etc.). **Our study suggests** significant sex differences in the intergenerational correlations in whole blood DNAm, particularly pronounced for ChrX, with more similarity between mothers and female newborns.

PrgmNr 2933 - Studying Regulatory Variation In Founder Populations To Identify Functional Rare Variants

View session detail

Author Block: **S. Ramdas**[1], B. Heil[1], R. Kember[1], M. Bucan[1], C. D. Brown[2]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ Pennsylvania, Philadelphia, PA
Disclosure Block: **S. Ramdas:** None.
With increasing sample sizes, GWAS and sequencing studies are now able to identify trait-associated variants with low allele frequencies. However, interpreting the function of these rare variants remains challenging. The functional interpretation of common disease-associated variation has been aided by the identification of variants regulating gene expression. However, expression quantitative locus (eQTL) studies are currently limited by their sample sizes, which make the functional interpretation of rare non-coding variants from these studies a challenge. In this study, we attempt an alternate approach for the functional interpretation of rare variants using a genetically isolated (founder) population. Founder populations harbor an increased burden of functional variants rare in more heterogeneous populations, making it possible to characterize their functional impact with smaller sample sizes. In this experimental design, we study genes with extremely high or low levels of expression (called â□□expression outliersâ□□). We posit that identifying expression outliers in a founder population will allow us to determine regulatory variants of large effect otherwise not detectable in commonly-studied populations.
The Amish represent a genetic isolate whose European ancestors settled in the Americas starting in the 17th century. We analyze RNA-seq data from
lymphoblastoid cell lines (LCL) obtained from 97 genotyped samples of a large multi-generational pedigree; this represents the first transcriptomic study on the Amish population. We identify 1,209 genes with eQTLs in this
cohort. While we see an eQTL replication rate of 72% in larger cohorts of LCLs, we find 206 eQTL genes unique to the Amish population at loci harboring genetic variation that is rare in large European cohorts. The unique pedigree structure of our cohort enables us to identify genomic segments shared identical by descent and proximal to gene expression outliers. We identify more than 6,000 outliers for gene expression, splicing, or allele-specific expression; more than 100 of which are associated with specific cis-regulatory haplotypes. We then added orthogonal epigenomic information to prioritize truly causal regulatory variants within these regions. These methods and results allow us to prioritize disease variants by identifying variants of large-effect, and lead to a more comprehensive annotation of the regulatory genome.

PrgmNr 2934 - Understanding the interplay of pancreatic cancer GWAS risk loci and cellular stress

View session detail

Author Block: **K. E. Connelly**, J. Zhong, I. Collins, Pancreatic Cancer Cohort Consortium, Pancreatic Cancer Case-Control Consortium, H. Arda, L. Amundadottir; NIH, Bethesda, MD
Disclosure Block: **K.E. Connelly:** None.
Pancreatic ductal adenocarcinoma (PDAC), the third leading cause of cancer-related deaths, is often difficult to diagnose. Genome-wide association studies (GWAS) have identified common susceptibility variants at twenty risk loci for PDAC. At numerous PDAC risk loci, the most likely functional genes are associated with recovery from stress-induced inflammation and acinar to ductal cell de-differentiation *in vivo*. We hypothesize that cellular stress is a crucial component mediating risk at non-coding pancreatic cancer susceptibility loci. Unfortunately, the currently available model systems used for functional characterization of pancreatic cancer GWAS risk loci do not mimic the stress and inflammation environment known to be associated with epidemiological PDAC risk factors. Our work aims to establish chemically induced conditions *in vitro* that mimic cellular stresses (i.e., endoplasmic reticulum (ER), oxidative, inflammation) to better understand GWAS risk loci in a system that better reflects conditions conducive to carcinogenesis. In a pilot study using the immortalized human pancreatic duct epithelial cell line HPDE, we examined the effects of various stress conditions on the likely functional gene at the chr7p14.1 (rs12701838, P= 3.9×10^{-9}) PDAC risk locus. This risk variants map to the final intron of *SUGCT*. However, expression quantitative trait locus (eQTL) analysis and our chromatin capture data in PDAC cell lines suggest *INHBA*, located approximately 1MB downstream, is the likely functional gene. We observed a significant decrease in *INHBA* expression in HPDE cells by reverse transcription quantitative PCR (RT-qPCR) under ER stress conditions (tunicamycin). Additionally, we observe a significant dose dependent increase in *INHBA* expression when cells are exposed to exogenous tumor necrosis factor alpha (TNFÎ±) as an inflammation stimulus. Further studies are ongoing to better understand the role of *INHBA* and the GWAS risk variants under stress and the contribution(s) to PDAC risk. Moreover, using concentration and time course data from our pilot study, we aim to expand our study to examine the transcriptional and chromatin changes under the various stress conditions in a global manner to establish a resource for additional current and future GWAS risk loci.

PrgmNr 2935 - Widespread sex differences in placental DNA methylation

View session detail

Author Block: **M. Ouidir**, S. Chatterjee, K. L. Grantz, F. Tekola-Ayele; Epidemiology Branch, Div. of Intramural Population Hlth.Res., Eunice Kennedy Shriver Natl. Inst. of Child Hlth.and Human Dev., NIH, Bethesda, MD
Disclosure Block: **M. Ouidir:** None.
Sex differences are common in fetal development, adverse pregnancy outcomes, and subsequent health. Accumulating evidence suggests that sex-biased epigenetic mechanisms in placenta may have a key role in the early life, but data are limited. We investigated sex differences in placental DNA methylation among 301 women from the NICHD Fetal Growth Studies - Singleton cohort.
We tested association between placental genome-wide methylation (Illumina 450k) and fetal sex using linear regression, adjusting for maternal race/ethnicity, gestational age at delivery, the first ten genotype principal components (PCs), methylation sample plate, the first three methylation PCs and putative cell-mixture estimated using surrogate variable analysis components (n = 20).
We identified 5124 differentially methylated CpGs in placenta between males and females (FDR P FOXN3; LogFC = 0.63, P = 1.5e-110) and that in females was found for cg11643285 (*RFTN1*; LogFC = 0.29, P = 1.7e-69). The genes (n = 2067) annotating the differentially methylated CpGs were enriched in gene ontology biological processes such as cornification (adjusted-P = 2.2e-10) and epithelium development (adjusted-P = 2.2e-10) as well as hallmark gene sets such as estrogen response (adjusted-P = 2.8e-3) and myogenesis (adjusted-P = 2.8e-3). In the EWAS Atlas the differentially methylated CpGs overlapped with CpGs previously associated with prenatal environmental exposures, pregnancy complications (e.g., preeclampsia, gestational diabetes mellitus, preterm birth), fetal and childhood growth and development-related traits, and adult complex diseases; the top three phenotypes with the largest number of overlapping CpGs were aging (n = 528 CpGs), gender (n = 476), and smoking (n = 366).
In all, our findings highlight that sex-biased methylation is widespread in the placenta. Accounting for fetal sex in placental studies can facilitate development of molecular diagnostics and interventions for diseases that exhibit disparities by sex.

PrgmNr 2936 - Wnt sensitive regulatory elements in neural progenitors harbor neuropsychiatric disorder and brain structure heritability

View session detail

Author Block: **N. Matoba**[1], B. D. Le[1], J. Valone[1], J. M. Wolter[1], D. Liang[1], J. Mory[1], M. L. Bond[1], J. L. Stein[2]; [1]UNC at Chapel Hill, Chapel Hill, NC, [2]UNC Chapel Hill, Chapel Hill, NC
Disclosure Block: **N. Matoba:** None.
GWAS for neuropsychiatric disorders and brain related traits show enrichment of heritability in tissue-, and cell-type specific regulatory elements (RE)s. These REs are often measured in post-mortem tissue so REs that are responsive to specific stimuli and involved in trait variation, are likely missed. External stimuli, like signaling factors, lead to transcription factor (TF) translocation to the nucleus and binding to DNA which alters RE activity. Previous studies have found that genetic loci associated with risk for neuropsychiatric disorders and interindividual differences in brain structure are enriched in genes of the Wnt pathway, a diffusible signaling molecule involved in brain development and patterning.
Here, we performed ATAC-seq on primary human neuronal progenitors from 93 donors treated with 5 nM Wnt3a or vehicle control. Of 173,034 accessible chromatin peaks identified, 38.6% of peaks were significantly differentially accessible in cells treated by Wnt3a as compared to vehicle conditions ((|LFC|>0; FDR As expected, we found that Tcf7 and LEF1, the known effectors of the Wnt signaling pathway, were the most enriched motifs identified in Wnt3a responsive REs (BH-corrected P = 1.58 Ã 10^{-32}; 8.65 Ã 10^{-32}, respectively), providing strong support that chromatin accessibility is a proxy for TF binding in response to external stimuli.
We next estimated whether genetic variation associated with neuropsychiatric disorders and brain traits are enriched in Wnt responsive REs. Using a partitioned heritability approach, we found that Wnt responsive REs showed significant heritability enrichment for schizophrenia, neuroticism, intelligence, cortical surface area (FDRP = 0.034), consistent with the spatial gradient of Wnt ligand secreted from the medially located cortical hem. These results demonstrate that genetic elements responding to Wnt signaling are significant contributors to multiple brain traits and act preferentially in brain regions expected to have high Wnt signaling activity.
In this study, we show that stimulus responsive regulatory elements play important roles in neuropsychiatric disorders and brain traits. In the future, analyzing individual variants associated with chromatin accessibility in response to stimuli may help explain individual GWAS loci (response-QTL) and be applied for pharmacogenomics approaches to optimize drugs based on genotype.

PrgmNr 2937 - Accurate identification of circRNA landscape and microRNA network in human oligodendroglia differentiation

View session detail

Author Block: **Y. Li**[1], F. Wang[1], P. Teng[2], L. Ku[2], L. Chen[3], Y. Feng[2], B. Yao[1]; [1]Dept. of Human Genetics, Emory Univ. Sch. of Med., ATLANTA, GA, [2]Dept. of Pharmacology and Chemical Biology, Emory Univ. Sch. of Med., ATLANTA, GA, [3]Dept. of Biostatistics and Hlth.Data Sci., Ctr. for Computational Biology and Bioinformatics, Indiana Univ. Sch. of Med., Indianapolis, IN
Disclosure Block: **Y. Li:** None.
Circular RNAs (circRNA), a novel class of covalently closed and poorly conserved non-coding RNAs derived from pre-mRNA back splicing that play complex roles in epigenetic gene regulation, are highly enriched in human brain and dysregulated in various brain disorders. Although numerous neuronal circRNAs have been identified, circRNAs in oligodendroglia (OL), the type of cells solely responsible for myelin formation on neuronal axons that underlie rapid long-range communication and affected in many myelin disorders, remain unexplored. This is partly due to difficulties in obtaining sufficient amount of human OL cells as well as the need for improved experimental approaches and computational tools to identify confident circRNAs and their precise sequence components. We effectively enriched circRNA using Poly(A)-tailing coupled with RNase R treatment (A-tailing) and developed a comprehensive computational framework for accurate circRNA identification and quantification, termed CARP (CircRNA identification using A-tailing RNase R approach and Pseudo-reference alignment). Using these approaches, we obtained highly confident, full-length circRNA sequence and distinct isoforms with identical back-splice sites. We also established conditions for inducing differentiation of the HOG human OL cell line, which harbors transcriptomic composition recapitulating OL progenitor cells in the human fetal brain. Using CARP, we identified OL-specific circRNA landscapes dynamically regulated during differentiation, which are distinct from human neuronal circRNAs. The production of a subset of circRNAs are independent of their parental transcripts and are predicted to be regulated by RNA editing in cis-regulatory elements such as A-to-I editing in *Alu* elements or trans-factors such as RNA-binding proteins. Interestingly, we found a substantial amount of alternative circularization of circRNAs forming clusters within the same loci, which apparently could exert additive effects in sponging miRNA and/or RNA binding-proteins. Furthermore, we discovered circRNA-miRNA networks that are regulated during OL differentiation, which are predicted to target well-defined signaling pathways known to promoting OL differentiation. Together, our improved experimental and computational methods identified differentiation-regulated human OL-specific circRNA landscape and circRNA-miRNA networks that are predicted to advance human OL development.

PrgmNr 2938 - Integrative multi-omics analysis to predict gene regulatory networks from genetic risk variants to phenotypes of Alzheimerâ□□s disease and COVID-19

View session detail

Author Block: **S. Khullar**[1], D. Wang, Ph.D.[2]; [1]Univ. of Wisconsin - Madison, West Windsor, NJ, [2]Univ. of Wisconsin - Madison, Madison, WI
Disclosure Block: **S. Khullar:** None.
Background Genome-wide association studies have found many genetic risk variants associated with Alzheimer's disease (AD). However, how these risk variants affect deeper phenotypes such as disease progression and immune response remains elusive. Also, our understanding of cellular and molecular mechanisms from disease risk variants to various disease phenotypes is still limited. To address these problems, we performed integrated multi-omics analysis from genotype, transcriptomics, and epigenomics for revealing gene regulatory mechanisms from disease variants to AD phenotypes. Method The first step in our integrative analysis is to cluster gene co-expression networks and identify gene modules for various AD phenotypes given population gene expression data. Next, we predict the transcription factors (TFs) that significantly regulate the genes in each module, as well as the AD SNPs interrupting the TF binding sites on the regulatory elements. Finally, we construct a full gene regulatory network linking SNPs, interrupted TFs, and regulatory elements to target genes for each phenotype. This network thus provides mechanistic insights of gene regulation from disease risk variants to AD phenotypes. Results We applied our analysis to population data and predicted gene regulatory networks for three major AD-relevant brain regions: hippocampus, dorsolateral prefrontal cortex (DLPFC), and lateral temporal lobe (LTL). These region networks provide a comprehensive functional genomic map linking AD SNPs to TFs and regulatory elements to target genes for various AD phenotypes. Comparative network analyses further revealed cross-region-conserved and region-specific regulatory networks. For instance, AD SNPs rs13404184 and rs61068452 disrupt the bindings of TF SPI1 that regulates AD gene INPP5D in the hippocampus and lateral temporal lobe. However, SNP rs117863556 strongly interrupts the bindings of TF REST to regulate GAB2 in the DLPFC only. Furthermore, driven by recent discoveries between AD and Covid-19, we also found that many genes from our networks regulating Covid-19 pathways are also significantly differentially expressed in severe Covid patients (ICU), suggesting potential regulatory connections between AD and Covid. Thus, we trained a machine learning model to predict severe Covid and prioritized the highly predictive genes as a set of AD-Covid marker genes. Thus, our results provide a deeper understanding of the interplay among multi-omics, brain regions and AD phenotypes including disease progression and Covid response.

PrgmNr 2939 - Learning interpretable cellular and gene signature embeddings from single-cell transcriptomic data

View session detail

Author Block: **Y. Li**[1], Y. Zhao[1], H. Cai[2], Z. Zhang[3], J. Tang[4]; [1]McGill Univ., MONTRÃAL, QC, Canada, [2]Peking Univ., Beijing, China, [3]Fudan Univ., Shanghai, China, [4]HEC, MONTRÃAL, QC, Canada
Disclosure Block: **Y. Li:** None.

The advent of single-cell RNA sequencing (scRNA-seq) technologies has revolutionized transcriptomic studies. However, large-scale integrative analysis of scRNA-seq data remains a challenge due to unwanted batch effects and limited transferabilty, interpretability, and scalability of the existing methods.

We present single-cell Embedded Topic Model (scETM). Our key contribution is the utilization of a transferable neural-network-based encoder while having an interpretable linear decoder via a matrix tri-factorization. scETM simultaneously learns an encoder network to infer cell type mixture and a set of highly interpretable gene embeddings, topic embeddings, and batch effect linear intercepts from multiple scRNA-seq datasets. We demonstrate scETM on 6 scRNA-seq datasets and compared it with 7 state-of-the-art (SOTA) scRNA-seq modelling methods. In terms of clustering known cell types, scETM achieved consistently competitive ARI, NMI, and kBET scores compared to the SOTA methods. Batch-overcorrection analysis shows that scETM does not overcorrect batch effects while learning embedding preserving the biological signals in contrast to some of more aggressive batch-correction methods.

Furthermore, scETM confers remarkable cross-tissue and cross-species zero-shot transfer-learning performance achieving the highest ARI across all 6 transfer-learning tasks. For instance, scETM trained on Tabula Muris (TM) on heterogeneous tissues clusters quite well the Mouse Pancreas (MP) cells (ARI: 0.941; kBET: 0.339); scETM trained only on the MP dataset cluster reasonably well the TM cells from diverse primary tissues, implying that it does not merely learn cell-type-specific signatures but also the underlying transcriptional programs that generalize to unseen tissues. Transfer learning between human and mouse pancreas are also accomplished well, implying its ability to capture conserved functions. Transfer learning between mouse and human primary motor areas presents a challenge due to the evolutionarily divergent functions of brains. Nonetheless, scETM conferred a much higher ARI of 0.696 for the MusMOp to HumM1C transfer-learning tasks and ARI 0.167 for the HumM1C to MusMOp task compared to the SOTA methods.

By exploring the gene and topic embedding via UMAP, we find that scETM-learned topics are enriched in biologically meaningful and disease-related pathways for Alzheimerâ□□s disease and Major depressive disorder. Lastly, scETM enables the incorporation of known gene sets into the gene embeddings, thereby directly learning the associations between pathways and topics via the topic embeddings. A preprint is available at bioRxiv.

PrgmNr 2940 - ModTools: a computational toolbox for rapid detection of DNA modifications and replications using Nanopore sequencing

View session detail

Author Block: **Q. Liu**[1], D. Georgieva[2], D. Egli[2], K. Wang[1]; [1]The Childrenâ□□s Hosp. of Philadelphia, Philadelphia, PA, [2]Columbia Univ., New York, NY
Disclosure Block: **Q. Liu:** None.
Background: DNA base modifications play critical roles in gene regulation and genome function. Existing bisulfate-conversion or antibody-based techniques with short-read sequencing have inherent limitations to detect DNA modifications. Nanopore sequencing enable genome-scale modification detection, but the performance and running speed need substantial improvement: they take over one day to analyze even just one PromethION data set and often encounter errors to stop prematurely. Additionally, genome-scale detection of synthetically introduced modifications enables DNA replication analysis by Replipore sequencing, but computational methods are under-developed.
Methods: We develop ModTools with three main modules implemented in multi-threaded C++ . The first module is comparison-based modification detection (improved NanoMod): modified bases are identified by comparing Nanopore signals between paired samples with and without modified bases. The second module is model-based modification prediction (improved DeepMod): a deep-learning framework is first trained for a specific type of modification (such as methylation), and predictions are made on new samples. The third module is to detect DNA replication origins based on DNA base analogs (such as BrdU or IdU) during DNA replication: the peak of predicted DNA replication origins from multiple reads are summarized from this module.
Results: We evaluated ModTools on several Nanopore data sets with different Thymidine analogs at known positions, and found that ModTools is one order of magnitude faster than NanoMod with improved accuracy. ModToolsâ□□ deep-learning module is evaluated to detect 5mC (5-methylcytosine) on E. Coli and human genomes, and Thymidine analogs on several datasets of yeast genomes, and again ModTools is one order of magnitude faster and more accurate. Applied on the replication data of yeast, ModTools can precisely identify DNA replication origins: ~90% peaks identified by ModTools are known DNA replication origins.
Conclusion: ModTools can detect both DNA modifications and DNA replications from Nanopore sequencing data. ModTools is faster and more accurate than previous tools and can be scaled to the analysis of PromethION data, and thus have great potential to speedup genome-scale analysis of DNA modifications and replications.

PrgmNr 2941 - *De novo* mucins in mammals: How parallel evolution of exonic repeats leads to new function

View session detail

Author Block: **O. Gokcumen**[1], P. Pajic[1], S. Shichen[1], J. Qu[1], S. Knox[2], S. Ruhl[1]; [1]Univ. at Buffalo, Buffalo, NY, [2]Univ. of California, San Fransisco, San Fransisco, CA
Disclosure Block: **O. Gokcumen:** None.
Evolution of novel gene function is a fundamental question in biology. We hypothesize that exonic repeats and their copy number variation contribute substantially to de novo evolution of new gene functions. Mucin proteins are densely glycosylated molecules critical for mucus layer formation, cellular signaling, and microbial interactions both on epithelial surfaces and in mucous secretions. Functionally, mucins share proline-, threonine-, and serine-rich (PTS) repeats that serve as O-glycosylation sites. However, mucin proteins do not belong to a homologous gene family. Instead, several mucin proteins, although biochemically and functionally related, have evolved independently from different ancestral genes. The mechanisms through which these mucins evolved provide a fertile ground to study parallel evolution of novel gene functions. To understand the evolution of mucin function, we chose to study mucins in saliva, not only because they are readily accessible for noninvasive sampling, but also because our previous studies had shown that these molecules are highly variable, presumably due to dietary or pathogenic selective pressures. To this end, we performed a bioinformatics analysis of 49 mammalian genomes to identify 27 hitherto undescribed mucins, highlighting 15 instances of independent, lineage-specific mucin evolution. By integrating, bioinformatic, phylogenetic, proteomic, and immunohistochemistry approaches, we documented multiple instances of likely evolutionary convergence, where these novel mucins have recurrently gained highly glycosylated repeat domains along with salivary expression in a lineage-specific manner. Collectively, our results identified the secretory calcium-binding phosphoprotein (SCPP) gene locus as a hotspot for mucin formation in mammals, where the secretory and proline-rich nature of proteins encoded by genes in this locus provides fodder for recurrent gain of mucin function. Our results have broad implications for understanding the role of exonic repeats in the parallel evolution of new gene functions, especially those involving protein glycosylation.

PrgmNr 2942 - A spatially aware likelihood test to detect sweeps from haplotype distributions

View session detail

Author Block: **Z. A. Szpiech**[1], M. DeGiorgio[2]; [1]Pennsylvania State Univ., University Park, PA, [2]Florida Atlantic Univ., Boca Raton, FL
Disclosure Block: **Z.A. Szpiech:** None.

The inference of positive selection in genomes is a problem of great interest in evolutionary genomics. By identifying putative regions of the genome that contain adaptive mutations, we are able to learn about the biology of organisms and their evolutionary history. One conspicuous genomic pattern that results from a recent or ongoing sweep is reduced genetic diversity and the presence of a single haplotype at high frequency in the case of a â□□hardâ□□ sweep or two or more haplotypes at high frequency in the case of a â□□softâ□□ sweep. To capture these patterns, we develop a composite likelihood method that identifies recently completed or ongoing positive selection by searching for extreme distortions in the spatial distribution of the haplotype frequency spectrum. Furthermore, the method simultaneously infers two parameters of the sweep: the number of sweeping haplotypes and the 'width' of the sweep, which is related to the strength of selection. We demonstrate that this method outperforms the leading haplotype-based selection statistics. Then, as a positive control, we apply it to two well-studied human populations from the 1000 Genomes Project and examine the haplotype frequency patterns at the *LCT* and *MHC* loci. To facilitate use of this method, we have implemented it in user-friendly open-source software.

PrgmNr 2943 - A systematic evaluation of phasing algorithms for downstream analysis in ancestry testing

View session detail

Author Block: **A. Rojas-Munoz**[1], G. Gonzalez-Arroyo[1], A. Lopez Pineda[2,1], K. Numakura[3], Y. Matsuda[4], N. Katsanis[1,1], T. Takano[3], C. D. Bustamante[5,1,6,7], A. Ioannidis[8,1]; [1]Galatea Bio, Hialeah, FL, [2]Amphora Hlth., Morelia, Mexico, [3]Awakens, Tokyo, Japan, [4]AWAKENS, Inc, Berkeley, CA, [5]Stanford Univ. Sch. of Med., Stanford, CA, [6]Univ. of Miami Sch. of Business, Miami, FL, [7]Chan Zuckerberg Biohub, San Francisco, CA, [8]Stanford Univ., Palo Alto, CA

Disclosure Block: **A. Rojas-Munoz:** None.

Background: In genetic ancestry testing, pipelines typically include raw data preprocessing, filtering, imputation of missing data, determination of haplotype phase, and the use of a genetic ancestry classification algorithm. Most phasing methods focus on improving metrics on test cases, but neglect to quantify performance for real world downstream applications. In particular, the quality of the phasing step for ancestry determination is typically not assessed. **Methods:** In this study, we have tested three state-of-the-art phasing algorithms for their ability to accurately infer local ancestry on data from a direct-to-consumer genetic testing research cohort. We compared Beagle (version 5.2), SHAPEIT (version 4), and Eagle (version 2) for their ability to effectively determine haplotype phase from both genotype array and NGS data. We evaluated phasing performance on the unphased version of the 1000 genomes dataset (version 3), using four metrics: 1) running time, 2) switch error rate as determined by comparison to the phased data, 3) phasing accuracy using the experimentally validated phasing data from individual N17828, and 4) local ancestry inference, where the ground truth was considered to be the continental-level populations. **Results:** We tested running time using 16 cores with multiple sample sizes, where the fastest phasing algorithm was SHAPEIT4 with 1 sample (18 s), and Beagle with 2,504 samples (515 s). When using 4 cores SHAPEIT4 was faster than Beagle by 36% with 2,504 samples (1119 s vs 722 s). Beagle had the lowest average error rate for 2,504 samples, and the best performance with the N17828 sample. Finally, the global ancestry inference was equivalently accurate across algorithms, when calculating admixture across the genome from a direct-to-consumer preexisting genetic cohort. Our ongoing analysis will further investigate the effect phasing algorithms have on local ancestry inference. **Conclusions:** Phasing algorithms are an important step in genetic applications pipelines. Understanding the effect of these algorithms in ancestry inference will help us determine the best alternative to use in industry-grade applications. The exponential growth of available genomewide data, alongside with the growth of cloud computing costs and architectures, makes it necessary to identify the best fit-for-purpose solution. We found that experiment-specific requirements would sway the final user to implement one pipeline over the alternative ones depending on their priorities.

PrgmNr 2944 - African ancestry polygenic risk scores improve Alzheimer disease risk prediction in individuals of African Ancestry

View session detail

Author Block: **F. RAJABLI**[1], K. L. Hamilton-Nelson[1], N. Kushch[1], L. D. Adams[1], T. Starks[2], I. Martinez Soto[3], R. Laux[4], P. Whitehead[1], B. Kunkle[1], M. L. Cuccaro[1], J. M. Vance[1], C. Reitz[3], G. Byrd[2], J. L. Haines[4], A. .[5], G. W. Beecham[1], M. A. Pericak-Vance[1]; [1]Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Wake Forest Sch. of Med., Winston-Salem, NC, [3]Taub Inst. for Res. on the Aging Brain, Columbia Univ., New York, NY, [4]Dept. of Population & Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [5]Univ. of Pensylvania, Philadelphia, PA

Disclosure Block: **F. Rajabli:** None.

Background: Recent genome-wide association studies (GWAS) in Alzheimer disease (AD) identified 37 significant genetic loci. However, most of these loci individually contribute a small amount of risk and have limited power to predict disease. One way to increase the prediction power is to aggregate the effect of these genetic risk variants. Polygenic risk score (PRS) is a valuable tool that combines the effect of each risk variant to improve the risk prediction of an individual from their genetic profile. Cross-population PRS studies showed that the over-representation of European ancestral populations in genomic research creates limited PRS transferability to non-European ancestral groups. Thus, the limited generalizability of PRS reduces its prediction power and may intensify healthcare disparities. In this study, we assessed and compared the PRS prediction accuracy of AD in individuals of African Ancestry (AA) using summary statistics from AA and non-Hispanic White GWAS studies. **Methods:** We generated PRS on a test dataset of 339 AA individuals (AD cases=111, controls=228) that was not included in AA GWAS study. First, we selected 'NHW markers' from the Kunkle et al. 2019 (pvResults: After p-value thresholding and clumping, we obtained 21 â□□NHW markersâ□□ and 32 â□□AA markersâ□□. PRS derived from the â□□NHW markersâ□□ failed to predict AD in the AA group (OR=0.82;pv=0.66) with the AUC 0.52, whereas the PRS scores of â□□AA markersâ□□ were significantly associated with AD (OR=2.46; pv=5.9x10-7) with the AUC of 0.67. **Conclusions:** We evaluated PRS scores based on 'EA markers' and 'AA markers' and found that African-ancestry PRSs improved AD prediction in individuals of African Ancestry. Our results show the impact of ancestral effects in PRSâ□□related research. PRS will help to develop AD prediction models for early identification in hopes of applying treatments to delay or prevent disease onset. Enhanced diversity in genetic studies is an essential step to create a transferable PRS across populations that will improve screening and prevention strategies.

PrgmNr 2945 - Dissection of ethnic and socio-cultural substructure of the South Asian Indian population using genetic data

View session detail

Author Block: **M. Kapoor**[1], N. Gosalia[2], C. V. Van Hout[3], B. Ye[4], J. D. Backman[5], O. Babu[6], N. England[6], J. L. Rodriguez-Flores[7], S. Jamuar[8], S. S. Jamuar[9], C. G. Gonzaga-Jauregui[10], D. Liu[11], W. McLaren[12], N. Sarkar[6], A. Shah[13], K. Sharma[13], S. Tuna[14], Y. Turpaz[13], G. Tzoneva[1], J. Overton[15], J. G. Reid[5], G. Abecasis[1], A. Baras[5], J. Picker[13], A. R. Shuldiner[5]; [1]Regeneron Pharmaceuticals, Tarrytown, NY, [2]New York, NY, [3]Natl. Autonomous Univ. of Mexico, Juriquilla QuerÃ©taro, Mexico, [4]Regeneron Pharmaceuticals Inc., Tarrytown, NY, [5]Regeneron Genetics Ctr., Tarrytown, NY, [6]Global Gene Corp., Singapore, Singapore, [7]Regeneron Pharmaceuticals, Inc., TARRYTOWN, NY, [8]Global Gene Coprporation, Singapore, Singapore, [9]Boston, MA, [10]Univ. Natl. AutÃ³noma de MÃ©xico (UNAM), Wautoma, WI, [11]New York, [12]Global Gene Corp, Cambridge, United Kingdom, [13]Global Gene Corp. PTE., Ltd, Singapore, Singapore, [14]Hinxton, Cambridge, United Kingdom, [15]Tarrytown, NY
Disclosure Block: **M. Kapoor:** Salary/Employment; Regeneron Pharmaceuticals.
Introduction: India is a diverse country with complex population sub-structure primarily shaped by its geography, language and social stratification. Past genetic studies on South Asian Indian population focused on a few selected subgroups of ethnically diverse populations to explore genetic admixture and migration patterns within geographical regions. Here we report the largest effort to date to tease apart the linguistic and socio-cultural diversity of all major subgroups residing within different geographical regions of India. **Methodology:** Unrelated subjects (N = 15,154) were recruited as part of an ongoing collaboration between the Regeneron Genetics Center and Global Gene Corporation to study genetic diversity across India. Subjects were recruited predominantly from five large states (Rajasthan, Maharashtra, Uttar Pradesh, Gujarat and Karnataka). Genotyping was performed using Illumina GSA-24v2-0_A2 arrays and exome sequencing. Array data was used to compute principal components (PCs). Dimensions of 20 PCs were further reduced to two axes using the Uniform Manifold Approximation and Projection (UMAP) method by considering the 10 nearest neighbors.**Results:** PC projections identified two large subgroups that were majorly defined by language. For example, the populations using Hindi dialect (Rajasthan and Uttar Pradesh) grouped separately from populations speaking Marathi (Maharashtra) or Gujarati (Gujarat). UMAP projections further segregated finer substructure within these populations by punctuating cultural differences within each state. For example, we observed differences between Hindi speaking â□□Tyagiâ□□ subgroups from Rajasthan and Uttar Pradesh. Although these groups originated from â□□Brahminsâ□□, geographical boundaries restricted gene-flow within each group. Further, we identified fine genetic differences between Marathi speaking populations from Maharashtra that recently diverged due to religious differences among different subgroups. Finally, exome sequencing identified drifted predicted loss-of-function and disease-associated variants of interest within each subgroup **Conclusion:** Understanding genetic diversity and architecture of disease from the Indian subcontinent, the second largest population in the world, is critical to advancing Precision Medicine for South Asian Indians throughout the world. This is the largest study to date to tease apart the linguistic, cultural and geographical differences that shape genetic diversity in India.

PrgmNr 2947 - Frequency enrichment of functional coding variants in a French Canadian founder population and its implication for inflammatory bowel diseases

View session detail

Author Block: **C. BhÃ©rer**[1], J-C. Grenier[2], G. Boucher[3], D. Taliun[1], C. R. Stevens[4], C. Laprise[5], P. Goyette[3], H. Huang[6], M. J. Daly[7], J. Hussin[8], v. Mooser[9], J. D. Rioux[10]; [1]McGill Univ., Montreal, QC, Canada, [2]5000 rue BÃ©langer, MontrÃ©al, Canada, H1T 1C7, Montreal, QC, Canada, [3]montreal Heart Inst, Montreal, QC, Canada, [4]The Broad Inst., Cambridge, MA, [5]UQAC, Saguenay, QC, Canada, [6]Boston, MA, [7]Massachusetts Gen. Hosp., Boston, MA, [8]Montreal, Canada, [9]McGill Genome Ctr., Montreal, QC, Canada, [10]Montreal Heart Inst. | Res. Ctr., Montreal, QC, Canada

Disclosure Block: **C. BhÃ©rer:** None.

The genetic features of founder populations with recent bottlenecks, causing some deleterious variants to rise to higher frequencies, can enhance the power of rare variant association studies. French Canadians (FC) from Quebec represent a recent founder population characterized by rapid demographic and territorial expansion leading to regional founder effects. FC are known for their unique disease heritage comprising more than 30 Mendelian conditions with higher incidence and/or specific phenotypes or mutations. Yet, to date, a comprehensive characterization of functional variation in this population is still lacking. Here, using whole-exome sequencing data from 2,625 FC (patients with inflammatory bowel diseases (IBD), unaffected parents and controls), we describe the distribution and functional impact of coding variation in this founder population. We find that 13% of protein-coding variants present in our sample at an allele frequency (AF) between 0.002 and 0.01 are 2.5-100 times more frequent (Fisher-exact p-value 0.002 classified as pathogenic or likely pathogenic in ClinVar (p-value SLC12A6 c.2436+1delG, AF=0.0075 vs AF_NFE=8.8 x 10-5, 85-fold enrichment), Vitamin D-dependent rickets type 1 (*CYP27B1* p.Val88TrpfsTer71, AF=0.0034 vs AF_NFE=9.1 x 10-5, 37-fold enrichment) and Leigh Syndrome FC type (*LRPPRC* p.Ala354Val, AF=0.0029 vs AF_NFE=0.00011, 26-fold enrichment). Finally, we investigate whether rare protein-altering and -truncating variants, enriched in FC by the founder effect, contribute to the risk of IBD using trio and case/control cohorts. Our results and approach illustrate the value of exome sequencing in founder populations for understanding their genetic make-up and frequency of disease-causing mutations, which has important public health relevance.

PrgmNr 2948 - Gene duplication analysis reveals genome evolution and adaptation of *Taenia* species

View session detail

Author Block: **N. T. Grube**, G. H. Perry; Pennsylvania State Univ., State College, PA
Disclosure Block: **N.T. Grube:** None.
Gene duplication (GD) has been studied in many parasitic organisms and is an important evolutionary force contributing significantly to species evolution, facilitating increases in gene dosage effect and potential functional diversity of expanded gene classes. In fact, high rates of small scale GD and functional diversifications contributed to species divergence amongst infective tapeworms (*T. asiatica, T. saginata, and T. multiceps*), an epidemiologically important parasite group whose rapid diversification has led to evolutionary specializations during host-parasite co-evolution. However, GD remains poorly understood from a cross species comparative analysis. Cross species comparative GD analysis will assess how particular GD events contribute specifically to species pathogenesis in, for example, human vs non-human tapeworm variants. Here, we perform genome-wide analysis of GD in *Taenia* species using computational biological tools to assess predominant GD patterns and evolutionary contributions in human vs non-human specific lineages. We leverage six publicly available draft genomes of *Taenia* species maintained by wormbase.org, 4 specialized human genomes, *T. asiatica* (n=2), *T. solium* and *T. saginata*, and 2 non-human genomes, *T. multiceps* and *T. taeniaeformis*, for comparative whole genome sequence analysis. Genomes maintained undergo iterative genome and transcriptome updates providing valuable data for investigating GD in this parasitic species. We employ a variety of computational analysis to assess GD, gene divergence following duplication and contributions of GD to species pathogenesis in human vs non-human *Taenia* lineages.

PrgmNr 2949 - Genomic patterns of natural selection and introgression at the alcohol dehydrogenase gene region are associated with agriculture in ethnically diverse Africans

View session detail

Author Block: **M. McQuillan**[1], A. Ranciaro[1], M. Hansen[1], S. Fan[2], W. Beggs[1], G. Belay[3], D. Woldemeskel[3], S. A. Tishkoff[4]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Human Phenome Inst., Sch. of Life Sci., Fudan Univ., Shanghai, China, [3]Dept. of Biology, Addis Ababa Univ., Addis Ababa, Ethiopia, [4]Univ. of Pennsylvania Sch. of Med., Philadelphia, PA
Disclosure Block: **M. McQuillan:** None.
The alcohol dehydrogenase (ADH) family of genes encode enzymes that catalyze the metabolism of ethanol into the toxic intermediate, acetaldehyde. Nucleotide variation in ADH genes affects the catalytic properties of these enzymes and has been associated with a variety of complex human traits, including alcoholism and cancer. Some ADH variants, including the *ADH1B*48His* (rs1229984) mutation in the *ADH1B* gene, confer a reduced risk of alcoholism and are under strong positive selection in multiple human populations. The advent of Neolithic agriculture and the associated increase in fermented foods and beverages is hypothesized to have been a selective force acting on such variants. However, this hypothesis has not been tested in populations outside of Asia. Here, we use genome-wide selection scans to show that the ADH gene region is enriched for single nucleotide polymorphisms (SNPs)s showing strong signals of positive selection in multiple Afroasiatic-speaking, agriculturalist populations from Ethiopia, and that this signal is unique among Sub-Saharan Africans. We also observe strong selection signals at many putatively functional variants in nearby lipid metabolism genes, which may influence evolutionary dynamics at the ADH region. Finally, we show that the haplotypes carrying these selected variants were introduced into Northeast Africa from a West-Eurasian source within the last ~2000 years and may have experienced positive selection following admixture. Interestingly, these selection signals are not evident in nearby, genetically similar populations that practice a hunting and gathering or pastoralist subsistence lifestyle, even though they experienced non-African admixture at a similar time. Together, our results support the hypothesis that the emergence of agriculture has shaped patterns of selection at the ADH gene region, and enhance our understanding of how adaptations to diverse environments and subsistence strategies have influenced the African genomic landscape. Supported by NIH training grant T32ES019851, NIH grant R35GM134957-01, and ADA grant 1-19-VSN-02.

PrgmNr 2951 - Mutational biases in the SARS-CoV-2 virus genome

View session detail

Author Block: **F. Mostefai**[1,2], D. Fournelle[1,2], J. Pelletier[2,3], J-C. Grenier[2], R. Poujol[2], D. J. Hamelin[4], Ã. Caron[4,5], J. Hussin[6,2]; [1]dÃ©partement de biochimie et mÃ©decine molÃ©culaire, FacultÃ© de MÃ©decine, UniversitÃ© de MontrÃ©al, MontrÃ©al, QC, Canada, [2]Ctr. de recherche, Inst. de Cardiologie de MontrÃ©al, MontrÃ©al, QC, Canada, [3]dÃ©partement de biochimie et mÃ©decine molÃ©culaire, FacultÃ© de MÃ©decine, MontrÃ©al, QC, Canada, [4]CHU Sainte-Justine Res. Ctr., MontrÃ©al, QC, Canada, [5]Dept. of Pathology and Cellular Biology, MontrÃ©al, QC, Canada, [6]dÃ©partement de mÃ©decine, FacultÃ© de MÃ©decine, UniversitÃ© de MontrÃ©al, Montreal, QC, Canada

Disclosure Block: **F. Mostefai:** None.

The Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2) is the highly mobile virus that led to the current ongoing COVID-19 pandemic. SARS-CoV-2 is a complex RNA virus from which genetic variants of concern are emerging all over the world. However, one important question is how SARS-CoV-2 has adapted to its human host since the first outbreak. In this study we are tracking the mutational patterns of the SARS-CoV-2 virus between and within infected individuals. We performed viral profiling on virus genomic data from the GISAID and NCBI repositories to understand the full scale of the virus' mutational diversity and characterize potential adaptation to human hosts. We used a computational genomics approach to curate and identify variants from SARS-CoV-2 consensus and RNASeq sequences. Using the identified variants, we tracked the viral mutational changes between (inter-host) and within COVID patients (intra-host). In our preliminary analysis of the pandemic first wave, we show that there is (a) a clear mutation bias towards fixation of C>U and G>U substitutions; (b) a selective advantage of the C>U substitutions that were not driven by the consequence on the coding proteins; (c) an excess of low-frequency G to U intra-host mutations that do not translate into higher fixation rates; and (d) a clustering of mutated positions with complementary substitutions U>G and A>C. Interestingly, in an analysis of the pandemic second wave sequences, we observed a reduction of the C>U mutational bias observed during the first wave. This result potentially reflects an evolutionary limitation of the virus mutational landscape. These results will help inform surveillance strategies for tracking emerging viral strains. Additionally, our observations will improve our understanding of the molecular mechanisms behind the adaptation of SARS-CoV-2 in response to the hostâ□□s defense strategy to fight infection.

PrgmNr 2952 - Novel unannotated protein-coding genes are expressed in multiple tissues

View session detail

Author Block: **W. Phu**[1,2,3], V. Luria[4], A. Karger[5], J. K. Goodrich[1,3], G. Tiao[1,3], A. O'Donnell-Luria[1,2,3]; [1]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Div. of Genetics and Genomics, Boston Children's Hosp., Harvard Med. Sch., Boston, MA, [3]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [4]Harvard Med. Sch., Boston, MA, [5]IT-Res. Computing, Harvard Med. Sch., Boston, MA

Disclosure Block: **W. Phu:** None.

In recent years, many protein-coding genes that were systematically missed by standard genomic algorithms have been discovered experimentally by proteogenomics tools such as proteomics or RiboSeq, highlighting that the identification of protein-coding genes in the human genome is incomplete. While a few unannotated genes have been shown to function in the nervous system, skin, gastrointestinal tract and male germline, the function and the evolutionary age of most unannotated genes remain unknown. We hypothesized that unannotated genes may provide additional candidate variants for rare disease diagnosis and assist in identifying new disease-causing genes. First, we found that, while some unannotated genes appear in many species and are thus evolutionarily ancient, most unannotated genes appear in few species (even just one) and are evolutionarily young. Thus, most unannotated genes are evolutionarily novel genes that arise by non-copying mechanisms (e.g. de novo from genomic DNA, lncRNAs that become coding, overprinting). Not being copied, such genes have no sequence similarity to genes in other species, which explains why they were missed by standard annotation tools. Second, we evaluated the expression patterns of a curated set of over 2,000 unannotated genes identified in previous proteogenomics studies, utilizing over 17,000 samples from the GTEx v8 dataset containing expression data from 54 human tissues. We determined in which tissues the individual genes were expressed, identified ubiquitous and tissue-specific genes, and explored how expression specificity correlates with gene evolutionary age. Third, we compared the expression levels of unannotated genes to the expression of Ensembl-annotated genes and of a control set of intergenic open reading frame (ORF) sequences. We found that most genes (60% of unannotated genes, 70% of annotated genes) are broadly expressed in >20 tissues. A fraction of genes showed higher tissue specificity: 14% of unannotated genes and 9% of annotated genes were expressed in 5 or fewer tissues; 7% of unannotated genes and 0.04% of annotated genes were expressed in exactly one tissue. To determine which unannotated genes are important for human disease, we will combine our expression data (abundance and tissue-specificity) with analysis of the variation in these genes in reference population and rare disease datasets. Thus, our expression analysis is the first step in elucidating the biological impact of the large number of overlooked novel human protein-coding genes.

PrgmNr 2953 - Overcoming constraints on the detection of recessive selection in human genes from population frequency data

View session detail

Author Block: **D. Balick**[1], D. M. Jordan[2], S. Sunyaev[3], R. Do[4]; [1]Brigham and Women's Hosp., Harvard Med. Sch., BOSTON, MA, [2]Icahn Sch. of Med. at Mt Sinai, New York, NY, [3]Harvard Med. Sch. and Brigham & Women's Hosp., Boston, MA, [4]Icahn Sch. of Med., New York, NY
Disclosure Block: **D. Balick:** None.
Identifying genes that evolve under recessive natural selection is a longstanding goal of population genetics research with important applications to gene discovery for traits and diseases. Unlike model organisms, identification of diploid selection coefficients in humans requires inference from natural population data. We assessed a range of population genetic and conservation-based measures designed to evaluate per-gene selective constraints and found that, while they are highly sensitive to genes under heterozygous selection, they ubiquitously fail to detect genes evolving under recessive selection. Additionally, more sophisticated likelihood-based statistics designed to explicitly infer recessive selection similarly lack power for any human gene of realistic length given the current size of population samples. However, extensive simulations suggested that enrichment of recessive genes may be detectable in aggregate for gene sets, but that this is sensitively dependent on the fraction of the whole genome that is recessive, an unknown quantity in humans. We designed a method to analyze genes in aggregate to identify enrichment for recessive purifying selection that is informed by population genetics simulations with realistic demography. Applying this to empirical gene sets successfully produced validating enrichments for strong recessive selection in genes previously inferred to be under recessive selection in a consanguineous cohort and in genes involved in autosomal recessive monogenic disorders, including in a set as small as 23 genes associated with effectively lethal recessive diseases. We created srMLgenes (https://jordad05.u.hpc.mssm.edu/srmlgenes), a publicly accessible tool that visualizes gene set enrichment patterns and allows users to upload their own human gene sets to search for recessive selection.

PrgmNr 2954 - PLIGHT: A tool to assess privacy risk by inferring identifying characteristics from sparse, noisy genotypes

View session detail

Author Block: **P. Emani**, G. GÃ¼rsoy, A. Miranker, M. B. Gerstein; Yale Univ., New Haven, CT
Disclosure Block: **P. Emani:** None.

The leakage of identifying information in genetic and omics data has been established in many studies, with single nucleotide polymorphisms (SNPs) shown to carry a strong risk of reidentification for individuals and their genetic relatives. While the ability of thousands or hundreds of thousands of SNPs (especially rare ones) to identify individuals has been demonstrated, here we sought to measure the informativeness of even a sparse set of noisy, common SNPs from an individual, by putting the genotype-based privacy leakage from an individual on quantitative footing. We present a computational tool, PLIGHT (â□□Privacy Leakage by Inference across Genotypic HMM Trajectoriesâ□□), that employs a population-genetics-based Hidden Markov Model of recombination and mutation to find piecewise matches of a sparse query set of SNPs to a reference genotype panel. Given the ready availability of auxiliary sources of noisy genotype data - such as acquiring small samples of environmental DNA or learning about someone's Mendelian diseases and physical characteristics - inference on sparse data becomes a genuine concern. We explore cases where query individuals are either known to be in databases or not, and consider both simulated â□□mosaicsâ□□ of genotypes (consisting of one or more source individuals) and actual genotype data obtained from swabs of coffee cups used by a known individual. Our findings are as follows: (1) Even 10 common SNPs (minor allele frequency > 0.05) often are sufficient to identify individuals in conventional genomic databases. (2) We are able to identify first-order relatives (parents, children and siblings) of query individuals with 20-30 common SNPs. (3) We find some potential for leakage of phenotypic information, based on a simulated attack by combining polygenic risk scores (PRSs) of the piecewise genotypic matches. We also found, for simulated mosaics of two individuals, that 20 common SNPs were often sufficient to find the correct identities of both component individuals. Finally, applying PLIGHT to coffee-cup-derived SNPs, we find that our tool is able identify the individual (when present in the reference database) using as little as 30 SNPs; when not present in the reference database, we use the inferred matches to the 30-90 query SNPs to perform a small degree of imputation of unobserved query SNPs. Overall, the tool could be used to determine the value of selectively masking released SNPs, in a way that is agnostic to any explicit assumptions about underlying population membership or allele frequencies.

PrgmNr 2955 - RaPID-Query: Towards a Fast and Accurate Real-Time Genealogical Search

View session detail

Author Block: **Y. Wei**[1], A. Naseri[2], D. Zhi[2], S. Zhang[1]; [1]Univ. of Central Florida, Orlando, FL, [2]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX
Disclosure Block: **Y. Wei:** None.
Thanks to the growing sizes of genetic databases, genetic genealogical search has become a potential effective approach to help individuals finding missing family members or law enforcement agencies locating suspects. While the problem of n-vs-n Identity-By-Descent (IBD) segment calling is well-studied, no efficient and accurate method exists for the problem of 1-vs-n IBD segment calling. Current approaches are inflexible and time consuming as they lack the ability to extract the accurate IBD information from the database only related to the target individual. Here, a new method, Random projection-based IBD detection (RaPID) query, RaPID-Query, is introduced that identifies IBD segments above a certain target length between a query individual and a population panel. An extended PBWT-Query algorithm was developed that guarantees the quick detection of IBD segments, while the random projection and merge algorithm allows the tolerance of genotyping errors, mutations, or gene conversions in detected IBD segments. The trimming algorithm keeps the accuracy of IBD segment boundaries. The detection power and accuracy of 1-vs-n RaPID-Query being comparable to the state-of-the-art n-vs-n IBD detection methods are demonstrated using simulated data using msprime with realistic population genetics models. For benchmarking efficiency, it is showed that RaPID-Query takes only 3.05 seconds on a single core on average to query an individual haplotype against the entire UK biobank of ~1 million haplotypes for all 22 autosomal chromosomes. Compared to the PBWT-Query without tolerating mismatches, RaPID-Query showed a better separation of the familial relatedness. It is anticipated that, since RaPID-Query can be easily parallelized, a fast and accurate real-time genealogical search for millions of samples would be possible.

PrgmNr 2956 - A Decision Science Approach to Understanding Motives that Underly Interest in Genomic Sequencing Results

View session detail

Author Block: **B. B. Biesecker**; RTI Intl., Washington, DC
Disclosure Block: **B.B. Biesecker:** Grant/Contracted Research Support (External); NIH and SCID Compass. Salary/Employment; RTI Employment. Consultant/Consulting Fees/Other Remuneration; Fees for thesis advising.

People often want genomic information and find it useful even in the absence of clinical utility. The science of decision making provides some possible explanations about how people use available information in the context of their values to make optimal choices. In particular, decision makers act on various motives. First, people have a desire to learn or understand, such as when the information provides self-knowledge (as genomic information does), when the information is â□□knowableâ□□ (i.e., if genomic sequencing has been done), and when their need for certainty is high. Second, learning information can help people feel more self-competent and self-efficacious because they learn personal information that they perceive as relevant to improving their health. Similarly, learning information can lead to a greater sense of control over personally relevant outcomes, which is strongly associated with better health. Fourth, seeking genomic information can also satisfy social motives - such as adhering to social norms and expectations regarding the seeking of personal health information - and can provide reassurance and other emotional benefits. Understanding these four types of motivations may help clinicians and investigators to better understand clientsâ□□ choices and the utility of making genomic information available to patients/consumers even when it does not have clinical utility. Expected benefits such as increased positive affect and wellbeing or increased self-efficacy and perceptions of control are no less real than more (seemingly) concrete clinical utility benefits. Notably, these motivations to learn results from genomic sequencing are also similar to motivations to learn other types of health-related information. In both contexts, motivations to learn self-information come from well-known psychological processes that underlie decision utility. Better integration of psychological and clinical science can enhance understanding of both.

PrgmNr 2957 - Addressing Barriers to Access in an Adult Kidney Genetics Clinic

View session detail

Author Block: **K. Bogyo**[1], N. Vena[1], J. Radhakrishnan[1], H. Milo Rasouly[2], A. G. Gharavi[3], J. Nestor[1]; [1]Columbia Univ. Med. Ctr., New York, NY, [2]Columbia Univ., New York, NY, [3]Columbia Univ, New York, NY

Disclosure Block: **K. Bogyo:** None.

Introduction: Nearly 1 in 10 patients with chronic kidney disease have a genetic etiology, yet utilization of genetic services is limited and disproportionally impacts those from medically underserved populations. To address the barriers to accessing such services, an adult genetics clinic was developed within the Division of Nephrology at a large academic medical center in an urban setting.

Methods: A bi-weekly adult Kidney Genetics Clinic was created, staffed by genetic counselors (GC) and nephrologists. In addition to in-person appointments, televisits were introduced due to the COVID-19 pandemic. Two visit types were available: Full Genetic Consults (staffed by GC and nephrologist) or Genetic Counseling visits (staffed by a GC). In parallel, genetic education initiatives were developed via bi-monthly genetic seminars, monthly sign-out rounds, a CME course, and GC student rotations.

Results: Between June 2019 and April 2021, the clinic received 247 referrals averaging 11 referrals/month, which increased to 23 referrals/month in 2021. Of those referred, 189 patients (76%) scheduled an appointment, 89% of which were via telehealth (21% by GC only) with a median wait time of 19 days from referral to appointment.

Most patients were in the NY tri-state area (NY, NJ, and CT), but 9% were from 9 other states. Medicaid was the primary insurance in 24% of patients. Genetic testing was ordered for 124 patients (66%), was not indicated in 5 patients (3%), and was already ordered prior to the visit in 12 patients (6%). In 5 patients, testing was not done due to financial concerns (4%). Categories of tests ordered included small gene panels of fewer than 50 genes (37%), large panels (49%), exomes (3%), and others, such as single variant testing or microarray (11%).

A diagnostic finding was identified in 27% of patients, with the highest yield in those with either cystic kidney disease or suspicion of Alport syndrome. Referrals to additional medical specialists were made in 67% of patients with a diagnostic finding and referrals were made to research studies in those with a non-diagnostic finding. **Conclusions:** The Kidney Genetics Clinic successfully utilized telehealth to provide genetic services to a broad population during the COVID-19 pandemic. A significant number of underserved patients were able to complete genetic testing and counseling. The high uptake of genetic services by patients with Medicaid and low appointment and test cancellation rates due to financial constraints suggest that cost was not a barrier to genetic testing in nephrology. Provider education was integral in creating a successful clinic and identifying patients who would benefit from genetic services.

PrgmNr 2958 - Addressing barriers to patient data sharing: Exploring the effects of employing electronic consent to obtain genetic testing records through ClinGen's Patient Data Sharing Program

View session detail

Author Block: **L. Paul**[1], J. M. Savatt[2], M. E. Good[1], A. R. Heidlebaugh[1], D. Azzariti[3], D. Jae[4], D. H. Ledbetter[5], B. McLarney[4], H. L. Rehm[6], E. R. Riggs[2], C. L. Martin[2]; [1]Autism & Dev.al Med. Inst., Geisinger, Lewisburg, PA, [2]Geisinger, Lewisburg, PA, [3]Broad Inst., Cambridge, MA, [4]Invitae, San Francisco, CA, [5]Geisinger, Danville, PA, [6]Massachusetts Gen. Hosp., Boston, MA
Disclosure Block: **L. Paul:** None.
The NIH-funded Clinical Genome Resource (ClinGen) is working to define the clinical relevance of genes and variants. Sharing genetic and health data is critical to the scalability and success of these efforts. In 2014, ClinGen launched the GenomeConnect registry to give patients who have had genetic testing the ability to share their de-identified data with publicly available databases, like NCBIâ□□s ClinVar. GenomeConnect also partners with other registry programs to continue to expand patient data sharing. Participants consent online, provide health history via surveys, and upload a copy of their genetic report to their account. They can choose to receive updates about their results and connect with researchers and other participants.

As of April 2021, 3,377 participants have enrolled in GenomeConnect, 30% of whom (n=1,016) have uploaded genetic reports. To date, GenomeConnect has submitted 1,419 variants to ClinVar. Uploading a genetic report is a known barrier to participation. In a pilot survey of 98 GenomeConnect participants, not having a copy of their report (27.6%), and not knowing how to upload it (30.6.%) were the main limitations. To address these barriers, ClinGen piloted an electronic consent (e-consent) process with the Invitae Patient Insights Network (Invitae PIN), one of our registry partners that engages individuals living with or at risk of developing a health condition. The e-consent process allowed ClinGen to request participantsâ□□ genetic reports directly from Invitae, rather than participants uploading reports. As of April 2021, 1,072 Invitae PIN participants consented to data sharing via ClinGen, and 623 (58%) had their genetic reports uploaded via e-consent. Combined with participant-uploaded reports, this process resulted in a total report upload rate of 76% (n=817/1,072), a significant increase over the 30% upload rate seen in GenomeConnect (pTo date, ClinGen has contributed 681 unique variants from 410 Invitae PIN participants to ClinVar. Of those variants, 17% (n=113) were novel to ClinVar and 71% (n=481) came directly from reports uploaded through e-consent. The submission contained detailed health and case-level data such as inheritance, when available.

Patient data sharing benefits patients and the genetics community, but difficulties uploading test results can limit participation. ClinGenâ□□s collaboration with the Invitae PIN offered participants an e-consent process that allowed for a higher percentage of reports shared, and more variants and associated phenotype data being submitted to ClinVar. These lessons can be applied to other interfaces to continue to enhance ClinGenâ□□s patient data sharing efforts.

PrgmNr 2959 - Benefits, harms and costs of newborn genetic screening for hypertrophic cardiomyopathy: estimates from the PreEMPT Model

View session detail

Author Block: **L. Galbraith**[1], K. D. Christensen[2], M. Gooch[1], R. J. Hylind[3,4], P. M. McMahon[1], G. O'Brien[3], N. K. Stout[1,4], J. M. Yeh[5], A. C. Wu[4,3,1]; [1]Harvard Pilgrim Hlth.Care Inst., Boston, MA, [2]Harvard Pilgrim Hlth.Care Inst. / Harvard Med. Sch., Boston, MA, [3]Boston Children's Hosp., Boston, MA, [4]Harvard Med. Sch., Boston, MA, [5]Harvard Med. Sch. & Boston Children's Hosp., Boston, MA
Disclosure Block: **L. Galbraith:** None.

Background

Hypertrophic cardiomyopathy (HCM) is a leading cause of sudden cardiac death (SCD) in children, but is traditionally considered an adult-onset condition. Genetic screening can identify newborns with pathogenic variants for HCM, but whether the benefits of population testing justify the harms and costs are unclear.

Methods

The Precision Medicine Prevention and Treatment (PreEMPT) Model is a simulation model that projects the benefits, harms, and costs of newborn genetic screening and usual care. We developed an HCM module to estimate diagnosed and undiagnosed HCM cases through age 20, with benefits accruing from surveillance of newborns with pathogenic variants. We estimated the prevalence of pathogenic variants in *MYPBC3, MYH7, TNNI3, MYL2, TNNT2, TPM1, MYL3*, and *ACTC1* using data from clinical studies, ClinVar, and gnomAD. Newborns with identified variants underwent guideline-based surveillance. Survival benefit was modeled via reductions in deaths from beta blocker treatment and placement of implantable cardioverter-defibrillators (ICDs) in symptomatic children. Outcomes examined included SCD, heart failure, and operative mortality; ICD placements and inappropriate shocks; and genetic screening, surveillance and treatment costs. Genetic screening costs were estimated at $40 per newborn.

Results

Among a cohort of 3.7 million US newborns, newborn genetic screening was projected to avert 44 of 188 HCM-related deaths (95% uncertainty interval [UI]: 10-123) before age 20 years, including 41 of 153 HCM-related SCDs (95%UI: 9 to 122), but increased costs by $279 million (95%UI: $271 million to $289 million). Genetic screening also resulted in 227 more children (95%UI: 70-556) receiving ICDs, 84 of whom (95%UI: 3-256) would experience inappropriate shocks by age 20. The estimated number total number of life years gained from newborn genetic screening was 1,065 (95%UI: 537 to 2,2719), and cost $261,720 per life year gained (95%UI: $102,830 to $900,450).

Conclusion

Newborn genetic screening could reduce mortality, but at costs above typical thresholds for cost-effectiveness. Findings suggest that population screening of newborns for HCM variants does not provide good value, and that targeted approaches of high-risk populations may be more appropriate.

PrgmNr 2960 - Caregiversâ understanding of 22q11 Deletion Syndrome and their communication with their affected adult children about the syndrome

View session detail

Author Block: **J. Rice**[1], J. Pagano[1], L. Lichten[1], K. B. Coleman[2], O. Y. Ousley[1], J. F. Cubells[3]; [1]Emory Univ., Altanta, GA, [2]Emory Univ., Atlanta, GA, [3]Emory Univ Sch Med, Atlanta, GA
Disclosure Block: **J. Rice:** None.
Background. 22q11 Deletion Syndrome (22q11DS) affects 1 in 4,000 live births with a median age of mortality of 41.5 years old. However, the majority of research on caregivers for patients with 22q11DS has focused on their experiences at the time of diagnosis, which usually occurs during childhood. We examined the knowledge of parents or caregivers of adults with 22q11DS about the genetic, medical, and psychiatric aspects of 22q11DS, and their communication patterns with their affected adult children about the diagnosis. **Methods**. Semi-structured interviews were conducted via video conference with caregivers (n = 8), recruited through the Emory Autism Center and the Emory 22q11DS database. Interviews were transcribed verbatim, then analyzed for themes by two independent coders using MAXQDA 2020. **Results**. Overall, caregivers were knowledgeable about the medical aspects of 22q11DS, such as cardiac anomalies and intellectual disability, but not the genetic aspects, including inheritance patterns and recurrence risk. Caregivers were knowledgeable about psychiatric manifestations, however many noted first learning of that association at the time of onset of psychiatric symptoms in their adult children. Frequency of communication was variable, ranging from never to regularly. Common topics discussed include reproductive decision making and taking ownership of medical care. The majority of caregivers indicated the importance of educating their affected adult children about the syndrome, although one caregiver did not feel this was important. Common barriers to communication included patient attitude, intellectual disability and psychiatric conditions. **Conclusions**. Our findings highlight the need for resources targeted towards educating caretakers of adults with 22q11DS about the genetic and psychiatric aspects of the condition, and a role for providers in facilitating communication between familial caretakers and their affected adult children.

PrgmNr 2961 - Clinical benefits, cost-effectiveness and value of further research for population-based newborn screening for Li-Fraumeni syndrome

View session detail

Author Block: **N. Kunst**[1], N. K. Stout[1], G. Oâ□□Brien[2], K. D. Christensen[1], P. M. McMahon[1], A. C. Wu[1], L. Diller[3], J. M. Yeh[4]; [1]Harvard Med. Sch. & Harvard Pilgrim Hlth.Care Inst., Boston, MA, [2]Boston Children's Hosp., Boston, MA, [3]Harvard Med. Sch. & Dana-Farber Cancer Inst., Boston, MA, [4]Harvard Med. Sch. & Boston Children's Hosp., Boston, MA
Disclosure Block: **N. Kunst:** None.
Background: Li-Fraumeni syndrome is an inherited familial predisposition to an elevated risk of developing cancer during infancy and childhood associated with pathogenic germline mutations in *TP53*. A national program for Li-Fraumeni syndrome newborn screening (LFS-NBS) can identify infants who might benefit from cancer surveillance and potential early detection, but it may also lead to unnecessary medical follow-up. We evaluated the potential clinical benefits, cost-effectiveness and the need for further research to implement LFS-NBS population-wide in the United States. **Methods:** We used the Precision Medicine Policy and Treatment (PreEMPT) model to simulate the health outcomes and costs associated with LFS-NBS. The model utilized data from the Surveillance, Epidemiology, and End Results Program, ClinVar and gnomAD and clinical studies. We simulated a birth cohort of newborns under usual care and LFS-NBS and evaluated diagnostic and clinical outcomes, life years (LYs) and costs associated with Li-Fraumeni syndrome-related pediatric cancers (adrenocortical carcinomas, choroid plexus carcinoma, osteosarcomas, rhabdomyosarcoma). We further applied decision-analytic methods to assess the value of additional data collection before implementing LFS-NBS population-wide, assuming a $100,000/LY willingness-to-pay threshold. **Results:** In a 4 million newborn cohort, LFS-NBS would identify 894 individuals with deleterious *TP53* variants who would then undergo surveillance; 67 of these (95%UI, 39-108) would develop Li-Fraumeni syndrome-related cancer before age 20. LFS-NBS leads to an overall 7.2% decrease (95%UI, 4.0-12.1%) in cancer-related deaths, and a 25.3% (95%UI, 8.7-49.9%) decrease in 5-year survivors at risk for radiation-related mortality. Furthermore, LFS-NBS was associated with incremental costs of $38 million and 361 LYs gained, corresponding to $106,009/LY. Analyses also indicated that LFS-NBS had a 40% probability of being cost-effective, suggesting considerable uncertainty, and identified research studies with positive expected benefits. More specifically, additional research on the probability of a deleterious germline *TP53* variant from 200 rhabdomyosarcoma cases would result in a gain of 349 LYs, which was close to the maximum expected benefit of further research to improve evidence on LFS-NBS (416 LYs). **Conclusions:** Our study indicates high uncertainty in the decision about population-wide LFS-NBS using current evidence, despite the potential of this strategy to improve health. We identified research studies that are feasible and inexpensive, and could help guide decisions about universal newborn screening.

PrgmNr 2962 - Communication about risk, testing, and test results with and through family members: A survey of families with Huntingtonâ□□s disease

View session detail

Author Block: **J. Bollinger**[1], S. Jose[2], K. Stuttgen[3], A. McCague[4], M. Terkowitz[5], C. Weston[6], D. J. Mathews[1]; [1]Berman Inst. of Bioethics, Johns Hopkins Univ., Baltimore, MD, [2]Hlth.Policy and Management, Johns Hopkins Univ., Baltimore, MD, [3]Dept. of BioMed. Res. Ethics, Mayo Clinic, Rochester, MN, [4]Policy and Program Analysis Branch, NHGRI/NIH, Bethesda, MD, [5]Intitute for Clinical and Translational Res., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [6]Bloomberg Sch. of Publ. Hlth., Baltimore, MD

Disclosure Block: **J. Bollinger:** None.

Huntingtonâ□□s disease (HD) is an autosomal dominant, progressive neurodegenerative disease for which there is currently no cure. Many people who are at risk for HD want to know if they will develop the disease, but others do not. For those who do get presymptomatic genetic testing (PGT), their results often shape their decisions, relationships, and lives. Communicating with family members about HD risk and genetic test results can strengthen relationships, but can also lead to tension and distance within families. To improve our understanding of how individuals and families communicate about and are shaped by HD risk and the results of HD testing, we developed an online survey for individuals aged 14 years and older who have HD in their family or who are the parent, spouse or partner of someone with a family history of HD. Methods: An online Qualtrics survey was developed based on prior qualitative research. The survey explored if, how, when, and by whom information about family history, genetic testing, and test results were communicated within the family and the impact of those communications on family relationships. We recruited both through direct invitations to 102 at-risk probands who enrolled in the presymptomatic genetic testing program for HD at Johns Hopkins between1986-98 (plus 15 family members who participated in noted prior research), including a unique, coded link to the online survey which they were encouraged to share with family members; and via broad-based outreach to the wider HD community through advertisements posted on numerous HD-related websites. Data were collected from 7/2020-6/2021. A total of 224 individuals completed the survey: 178 â□□probandsâ□□ (those with a family history of HD including themselves and/or family members) and 46 â□□partnersâ□□ (those who are/were in a romantic relationship with an individual with HD or a family history of HD). More than two-thirds of at-risk probands (68%) had pursued PGT and 61% of partners reported that their partners had pursued PGT. While a majority of probands shared their PGT results with family members, 25% reported deliberately withholding results, and 21% believed that family members are withholding information about their own genetic test results. Overall, 45% of respondents reported that their decision to be tested (or not) strengthened their family relationships and 25% reported increased conflict and distance. These survey data will help improve our understanding of how information about HD is communicated within and through families, how it influences the trajectories of at-risk individuals and their families, and how care teams might best support their decision-making and communication.

PrgmNr 2963 - Cost-efficient risk assessment for COVID-19 at work and play using machine learning

View session detail

Author Block: **K. Tretina**; Meenta Inc., Boston, MA
Disclosure Block: **K. Tretina:** None.
While many countries have taken action to slow down the spread of COVID-19, in the United States much of the burden of testing falls on private labs at the cost of businesses. To many of these businesses, the potentially immense costs of testing seem excessive. Here, we are developing an app that uses machine learning to prioritize the testing of quarantined employees as a mechanism for getting employees back to work quickly. This app leverages a large database of symptoms and test results to predict the risk of SARS-CoV-2 infection in symptomatic people with high accuracy in real time, allowing businesses to properly allocate resources and often saving them thousands of dollars in reduced testing costs and increased employee return rates.

PrgmNr 2964 - Deciding to pursue pediatric whole genome sequencing: Exploring value from parents' perspectives

View session detail

Author Block: **K. Dunsmore**[1,2], L. Gotlib Conn[3,4,2], M. Zubairi[5], V. Venkataramanan[1], E. Liston[6,7], R. Kim[8,7], R. Z. Hayeems[1,9,2]; [1]Child Hlth.Evaluative Sci., Hosp. for Sick Children, Toronto, ON, Canada, [2]Inst. of Hlth.Policy, Management and Evaluation, Univ. of Toronto, Toronto, ON, Canada, [3]Dept. of Anthropology, Univ. of Toronto, Toronto, ON, Canada, [4]Evaluative Clinical Sci., Sunnybrook Res. Inst., Toronto, ON, Canada, [5]Dept. of Pediatrics, Faculty of Hlth.Sci., McMaster Univ., Hamilton, ON, Canada, [6]Ted Rogers Ctr. for Heart Res., Hosp. for Sick Children, Toronto, ON, Canada, [7]Div. of Clinical and Metabolic Genetics, The Hosp. for Sick Children, Toronto, ON, Canada, [8]Ted Rogers Ctr. for Heart Res., The Hosp. for Sick Children, Toronto, ON, Canada, [9]Ted Rogers Ctr. for Heart Res., The Hosp. for Sick Children, Toronto, ON, Canada

Disclosure Block: **K. Dunsmore:** None.

Objective: Health technology assessment calls for careful consideration of patient and societal values. Existing research has demonstrated that personal and societal values influence how individuals and families orient and respond to genetic testing. Data that reflect on these values in the context of whole genome sequencing (WGS) are only beginning to emerge. The objective of this study was to understand how personal, familial, and societal factors influence parentsâ decisions to pursue WGS and receive secondary findings for their child.

Methods: Embedded in the Ted Rogersâ Cardiac Genome Clinic (CGC) at the Hospital for Sick Children (Toronto, ON), parents of children with congenital heart defects or cardiomyopathies who were offered WGS were eligible to participate. Prior to the receipt of WGS results, telephone-based semi-structured interviews explored personal, familial, and societal factors that influenced parentsâ decisions to pursue WGS and receive secondary findings. Interviews were audio-recorded and transcribed verbatim. Guided by the social ecological model and Kohlerâs construct of personal utility, deductive and inductive coding identified core themes.

Results: We conducted 19 interviews with English-speaking mothers of children who were offered WGS. During decision-making, respondents described the personal value they expected to gain from WGS. Gaining an understanding of their childâs current and future health was perceived to be a central benefit and acted as a precursor to enabling cognitive, emotional, and behavioural control. However, parentsâ enhanced sense of control required a readiness to learn information related to both primary and secondary findings. Parentsâ decisions were also impacted by social relationships and norms; WGS was valued as a mechanism for mitigating blame and legitimizing disability.

Conclusion: Findings from this study shed light on the acceptability of the technology from parentsâ perspectives and may inform strategies for education, counselling, and health technology assessment.

PrgmNr 2965 - Diagnostic sequencing to support genetically stratified medicine in a tertiary care setting

View session detail

Author Block: **L. Bier**[1], N. Lippa[1], A. Revah-Politi[1], H. J. May[2], S. Kushary[1], N. Vena[2], J. L. Giordano[2], H. Milo-Rasouly[2], E. Cocchi[2], T. Sands[2], R. J. Wapner[2], K. Anyane-Yeboa[3], A. G. Gharavi[2], D. B. Goldstein[2]; [1]Inst. for Genomic Med., Columbia Univ. Irving Med. Ctr., New York, NY, [2]Columbia Univ. Irving Med. Ctr., New York, NY, [3]Dept of Pediatrics, Columbia Univ. Irving Med. Ctr., New York, NY
Disclosure Block: **L. Bier:** None.
Purpose: The goal of stratified medicine is to identify subgroups of patients with similar disease mechanisms and specific responses to treatments. To prepare for stratified clinical trials, genome-wide genetic analysis should occur across clinical areas to identify undiagnosed genetic diseases and new genetic causes of disease. **Methods:** To advance genetically stratified medicine we have developed and implemented broad exome sequencing (ES) infrastructure and research protocols at Columbia University Irving Medical Center/NewYork-Presbyterian Hospital (CUIMC/NYPH). **Results:** We enrolled 4890 adult and pediatric probands and identified a primary result in 568 probands. The cohort was phenotypically and demographically heterogeneous as enrollment occurred across multiple specialty clinics (i.e. epilepsy, chronic kidney disease, fetal anomaly). New gene-disease associations and phenotypic expansions were discovered across clinical specialties. **Conclusions:** Our study processes have enabled the enrollment and exome sequencing/analysis of a phenotypically and demographically diverse cohort of patients within one tertiary care medical center. Since all genomic data is stored centrally with permission for longitudinal access to the electronic medical record, subjects can be re-contacted with updated genetic diagnoses or for participation in future genotype-based clinical trials. This infrastructure has allowed for the promotion of genetically stratified clinical trial readiness within the CUIMC/NYPH healthcare system.

PrgmNr 2966 - Digital tools for delivering genomic services: A systematic review

View session detail

Author Block: **D. Assamad**[1], W. Lee[1,2], S. Shickh[2,3], S. Luca[1], M. Clausen[3], C. Somerville[1], A. Tafler[1], A. Shaw[3], R. Hayeems[1,2], Y. Bombard[2,3]; [1]Child Hlth.Evaluative Sci., The Hosp. for Sick Children, Toronto, ON, Canada, [2]Inst. of Hlth.Policy, Management, and Evaluation, Univ. of Toronto, Toronto, ON, Canada, [3]Genomics Hlth.Services Res. Program, Li Ka Shing Knowledge Inst., St. Michaelâ□□s Hosp., Unity Hlth.Toronto, Toronto, ON, Canada

Disclosure Block: **D. Assamad:** None.

Background: Patient-facing digital tools including chatbots are increasingly being used in practice to supplement genetic counseling. To inform efforts to develop, evaluate and scale the use of effective tools, we conducted a systematic review to synthesize existing evidence on the methods to develop and evaluate these tools.

Methods: A search of peer-reviewed, empirical literature from Jan 2010 - Mar 2021 yielded 5072 citations. MeSH terms included: telemedicine, AI, digital, virtual, eHealth, genetic testing/counseling. The primary outcomes were patient-reported usability (e.g., satisfaction) and outcomes of tool use (e.g., knowledge, psychosocial well-being). Secondary outcomes included system-related (e.g., service efficiencies), provider-reported (e.g., workflow integration), and diversity, equity, and inclusion (DEI) dimensions.

Results: In total, 3368 abstracts were screened; 87 papers met inclusion criteria. Seventy-one distinct tools were identified. Their intended clinical settings were: cancer (48%), adult non-cancer (17%), prenatal (17%), pediatric (8%), primary care (5%) & other (9%). Most tools involved multiple steps in the genetic service delivery pathway including education (86%), decision making (51%), psychosocial needs or values assessment (30%), clinical assessment (24%), return of results (16%), post-test counselling & management (16%), consent (1%) & sequencing & interpretation (1%). However, none covered all components of the pathway. Patient reported usability was measured in 72% of studies; of which, 69% reported their tool was acceptable, 53% as satisfactory & 23% reported that users would recommend their tool. As a result of digital tool use, 84% of studies reported an improvement in at least one of the following patient outcomes: knowledge, psychosocial wellbeing, behavioural/management changes, family related communication, decision making or level of engagement. Many studies reported on DEI outcomes including education level (85%), ethnicity (77%) & income (37%). However, other outcomes were less represented such as literacy levels (general, health or digital) (18%), insurance coverage (11%), marital status (10%), & religion (5%).

Conclusion: Although numerous digital tools are available, none that cover the complete genetic service pathway were found. Gaps also remain with respect to DEI considerations and the limited range of clinical settings the tools are designed to support, such as primary care or pediatrics. Findings identify priorities and strategies for developing evidence-based digital tools in genomic medicine.

PrgmNr 2967 - Emerging public policy and perspectives on biometrics (including DNA and facial imaging) in the US

View session detail

Author Block: **J. K. Wagner**; Geisinger, Danville, PA
Disclosure Block: **J.K. Wagner:** Salary/Employment; Geisinger Health System.
Background: DNA and facial imaging have become increasingly important to precision medicine. They are also both being used more frequently and in expanded societal contexts as biometrics (i.e., measurable human characteristics used to recognize the identity or verify claimed identity of an individual). While extensive efforts have been taken by the scientific community to ensure adequate protection of the privacy and security of DNA data, less attention has been given to the challenges posed for responsible stewardship of facial images and derived data. With increased availability of facial images for precision health and media coverage of the perils of facial recognition technologies, a better understanding of biometrics policy development and public policy perspectives regarding biometric data protections is needed. ***Methods:*** Congress.gov and OpenStates.org were used to identify federal and state legislative activity during the most recent two years related to the collection, use, or management of biometrics using eight search strings: DNA (D); genetic data (GD); genetic information (GI); biometrics (B); biometric data (BD); biometric information (BI); faceprint (F); and facial recognition (FR). Bills identified with multiple search terms and identical bills were consolidated. Information was extracted from each identified bill to enable subsequent qualitative analysis. Survey responses from a diverse sample of U.S. adults were obtained through quota sampling via Qualtrics Panels in late 2020. Descriptive statistics were calculated. **Results:** A total of 229 unique federal bills were identified. No bill was identified by more than four separate search terms, which occurred via two combinations: (BD, B, GD, GI) and (BI, BD, B, GD). Surveys were completed by 4,048 respondents with diverse political views (25% liberal, 39% moderate, 27% conservative). Most respondents (60%) indicated the COVID-19 pandemic has not changed their opinion regarding use of biometrics in society, but nearly half (49%) reported their opinion of biometrics is less favorable now than it was five years ago. While most expressed concerns about possible misuse of their personal information generally (50% very and 38% somewhat concerned), nearly one-third (31%) indicated that biometric data protection and privacy laws in the US are adequate. ***Conclusions:*** These preliminary findings of emerging public policy and perspectives can be used to design in-depth ELSI research that will facilitate an advanced understanding of the implications of precision health initiatives and further refinement of reasonable and effective data management and sharing practices and policy.

PrgmNr 2968 - Ethical issues in pediatric gene therapy research

View session detail

Author Block: **A. Bateman-House**[1], L. Shah[2], C. Hunt[1], Pediatric Gene Therapy and Medical Ethics (PGTME)Working Group; [1]NYU Grossman Sch. of Med., New York, NY, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY
Disclosure Block: **A. Bateman-House:** Grant/Contracted Research Support (External); Johnson & Johnson, Pfizer. Consultant/Consulting Fees/Other Remuneration; Alexion.
There are myriad ethical challenges surrounding gene therapy clinical trials conducted on children. Some of these challenges, such as insuring appropriate practices surrounding the process of obtaining informed consent by a surrogate decision-maker and assent from a capable child participant, are common to all human subjects research in pediatric populations. Others, such as the current irreversibility of gene therapy interventions and participants' resultant inability to effectively withdraw from research, are more specific to gene therapy. And some are specific to pediatric gene therapy clinical trials: for example, the challenges of working with pediatric trial participants who transition to adulthood during the multi-year follow-up period. Furthermore, there are evergreen research ethics issues of equity of access and appropriate risk/benefit balancing. We identify some of the most pressing ethical challenges, to promote increased awareness and understanding of them, and share nascent best practices for clinical research in this rapidly-evolving area.

PrgmNr 2969 - Genes to Mental Health Network Report: Stakeholder perspectives on research participation

View session detail

Author Block: **T. Crowley**[1], L. White[1,2], B. M. Finucane[3], S. Garcia-Minaur[4], G. M. Repetto[5], M. Fischer[6], S. Jacquemont[7], M. van den Bree[8], R. Gur[1,2], A. Maillard[9], K. Donald[10], A. S. Bassett[11], A. Swillen[12], D. M. McDonald-McGinn[13]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Perelman Sch. of Med. of the Univ. of Pennsylvania, Philadelphia, PA, [3]Geisinger Hlth.System, Lewisburg, PA, [4]Hosp. Univ.rio La Paz, Madrid, Spain, [5]Clin Alemana- Univ Desarrollo, Santiago, RM, Chile, [6]Univ. of Rostock, Rostock, Germany, [7]CHU Sainte justine Res. Ctr., Montreal, QC, Canada, [8]Cardiff Univ., Cardiff, United Kingdom, [9]Lausanne Univ. Hosp., Lausanne, Switzerland, [10]Red Cross War Mem. Children's Hosp., Cape Town, South Africa, [11]Univ of Toronto, Toronto, ON, Canada, [12]KU Leuven, Leuven, Belgium, [13]Children S Hosp. of Philadelphia, Philadelphia, PA

Disclosure Block: **T. Crowley:** None.

Background: Feedback from research study participants is rarely collected. Therefore, investigators have limited understanding regarding their motivation to participate. Gaining insight is relevant to the 22q community. G2MH surveyed perceptions of eligible stakeholders, including motivation for initial and continuing study involvement and opinions on investigation priorities. **Methods**: A REDCap survey was built in English on instruments from Vanderbilt and EURORDIS, translated into 11 languages using DeepL software, reviewed by 6 native speakers, and distributed by 22 advocacy groups including 22q11.2 organizations. **Results**: Respondents included 1,035 people from 29 countries. 704 completed the entire survey. 82% were unaffected parents. 44% previously participated in research. 77% reported a chromosome 22q11.2 affiliation. Top reasons for initially partaking included compensation, provider encouragement, free healthcare, and positive previous experiences. Motives for leaving included treatment risks/pain/result non-disclosure. Participants stayed for access to care/information, helping others, improving QOL. Importantly, participants wanted summaries of results/labs and flexible schedules. **Conclusions**: This study provides invaluable insight for planning research studies in partnership with stakeholders. Payment was an initial motivator for joining, as was encouragement from a clinician or caregiver, but participants ranked non-monetary benefits as reasons to remain engaged. Notably, satisfaction was high overall for those previously participating in studies. Stakeholderâ□□s identified lack of public funding for rare disease research, lack of public awareness, difficulty in generating interest in participation, and a paucity of study subjects due to low prevalence as obstacles to supporting such investigations. G2MH hopes to address these issues by analyzing data across rare CNVs including 22q11.2.

PrgmNr 2970 - Hearing in Generation Genome: Comprehensive Newborn Hearing Screening through SEQaBOO (SEQuencing a Baby for an Optimal Outcome)

View session detail

Author Block: **C. C. Morton**[1,2,3], J. Shen[4], S. S. Amr[1,2], K. T. Booth[2], M. Chau[5], Y. Chekaluk[6], K. Choy[7], M. S. Cohen[8], Z. Dong[5], K. E. Gregory[1], J. Hochschild[9], L. J. Johnson[1], M. A. Kenna[10], A. E. Lorenzo[1], L. McGrath[1], C. O. Mitchell[1], J. Perry[10], A. E. Shearer[10], M. E. Stenerson[8], A. B. Giersch[1]; [1]Brigham & Women's Hosp, Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Univ. of Manchester, Manchester, United Kingdom, [4]Brigham & Women's Hosp., Harvard Med. Sch., Boston, MA, [5]The Chinese Univ. of Hong Kong, Hong Kong, China, [6]Brigham and Women's Hosp./Harvard Med. Sch., Brookline, MA, [7]Chinese Univ. of Hong Kong, Hong Kong, China, [8]Massachusetts Eye and Ear Infirmary, Boston, MA, [9]Harvard Univ., Cambridge, MA, [10]Boston Children's Hosp., Boston, MA

Disclosure Block: **C.C. Morton:** None.

Early genetic screening in congenitally deaf and hard-of-hearing (DHH) newborns or in non-penetrant DHH babies at birth is an integral component of a comprehensive newborn hearing screening (NBHS) program, which also includes testing for congenital CMV (cCMV) infection and physiologic hearing screening. Recently, compelling data have shown that genetic testing and physiologic screening in NBHS positively impact clinical outcomes and quality of life. SEQaBOO is a leading platform for assessing DHH newborns and evaluating parental attitudes concerning genomic medicine. Parents of newborns referred following a positive NBHS or at confirmatory diagnostic audiometry may enroll for SEQaBOO (comprehensive genome sequencing and variant interpretation, including optional ACMG secondary findings v3.0 for parents, plus annual surveys) or SEQaBOO surveys only. Annual surveys collect data on family medical history, health information and evolving attitudes on genomic medicine. Sequencing data and interpretation are possible prior to diagnostic audiometry and have influenced standard-of-care follow up; confirmatory genetic testing is required per SEQaBOO IRB protocol. Of 203 families approached, 73% (n=149) enrolled with >50% choosing to enroll in genome sequencing. Most NBHS referrals are unilateral and pass diagnostic audiometry with negative genetic results. Among SEQaBOO babies ultimately diagnosed as DHH (n=28), a genetic etiology of *GJB2* and *SLC26A4* variants was reported (n=6). For 13 babies, genetic diagnoses were inconclusive with only one pathogenic/likely pathogenic variant in a recessive gene with or without a second variant classified as VUS or a VUS in a gene associated with dominant DHH. SEQaBOO incorporates cCMV testing into genome sequencing analysis of babiesâ cord blood and confirmed known cases of cCMV (n=2). Genome wide CNV analysis revealed *STRC* and *OTOA* deletions from read count analyses. Additionally, our platform detected the incidental finding of a chromosomal translocation and can identify absence of heterozygosity. Nine cases have non-genetic etiologies. Three ACMG pathogenic secondary findings in adults were disclosed (*COL3A1*, *NF2* and *PKP2*). Feedback from parental surveys is positive with 61.5% of parents acknowledging the benefits of receiving genome sequencing results on themselves, 67.6% commenting on benefits to both themselves and their child and 62.6% stating that the genetic sequencing results have helped them understand their childâs DHH. Newborn screening has long been driven by technology and can now embrace integration of genetic tests to provide life-altering treatments and management to DHH infants.

PrgmNr 2972 - NHGRIâ□□s Inter-Society Coordinating Committee for Practitioner Education in Genomics: Multi-disciplinary genomics resource development

View session detail

Author Block: **D. Messersmith**[1], K. Jacoby Morris[1], S. Teixeira[1], R. L. Haspel[2,3]; [1]Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [2]Beth Israel Deaconess Med. Ctr., Boston, MA, [3]Harvard Med. Sch., Boston, MA
Disclosure Block: **D. Messersmith:** None.
Formed in 2013, the mission of the Inter-Society Coordinating Committee for Practitioner Education in Genomics (ISCC-PEG), hosted by the National Human Genome Research Institute, is to improve genomic literacy of healthcare providers by facilitating interactions among key stakeholders (genome.gov/iscc). Through ISCC-PEG, members collaborate to identify educational needs and potential solutions and develop resources to promote effective practice of clinical genomic medicine. There are currently 230 members representing over 130 organizations including companies, academic institutions, professional societies and government institutes. Members span the health professional spectrum including physicians, nurses, physician assistants, pharmacists, genetic counselors, and medical educators. Members share expertise and creative educational approaches to provide their colleagues who may not have genomics training with free teaching and self-study resources. Ongoing issues and work in progress are discussed via monthly project group calls, bi-monthly plenary calls, and during an annual meeting.

Members propose and initiate project groups to address specific areas in genomics education. The five current project groups focus on rare diseases, diversity and inclusion, direct-to-consumer genetic testing, obstetrics and gynecology, and pharmacogenomics. These multi-disciplinary teams have developed instructional tools, publications, professional society presentations, and social media campaigns. The ISCC-PEG Scholars program, initiated in 2020, pairs students with an interest in a genetics-related career with an ISCC-PEG member mentor to work on an educational project during a two-year term.

During June 7-11, 2021, 56 organizations collaborated with ISCC-PEG to disseminate healthcare provider genomics education resources through a social media campaign. The campaign hashtag #MedGeneEd21 provided a unifying theme in which contributors could tag their content. Through live engagement events, resource sharing and webpage promotion, the 5-day campaign was effective in disseminating resources and raising awareness with over 3 million impressions on Twitter.

Here we plan to present example resources, including self-study modules and case studies. We will also further describe the ISCC-PEG approach, data analytics of the social media campaign, and the positive impact of the committee to create and disseminate healthcare provider genomics education resources across healthcare communities.

PrgmNr 2973 - Polygenic risk scores change primary care providersâ preventive care of racially diverse patients: Results of a national survey with randomized case scenarios

View session detail

Author Block: **B. Kerman**[1], C. A. Brunette[2], A. A. Lemke[3], E. Harris[4], A. Antwi[5], N. Jones[2], J. L. Vassy[6]; [1]Harvard Med. Sch. at Brigham and Womens Hosp., Boston, MA, [2]VA Boston Hlth.care System, Boston, MA, [3]NorthShore Univ Hlth.System, Evanston, IL, [4]Harvard Med. Sch., Boston, MA, [5]VA Boston Hlth.care System, Boson, MA, [6]Harvard Med. Sch. at VA Boston Hlth.care System, Boston, MA
Disclosure Block: **B. Kerman:** None.
Background: Polygenic risk scores (PRS) may enable precision prevention if clinicians accept their use in medical decision-making, but their lower accuracy for non-European ancestry individuals poses equity concerns. To assess the impact of PRS on medical decision-making, and whether the impact differs by patient race, we conducted a national survey of U.S. primary care physicians (PCPs).
Methods: Between 4/18/21-5/13/21, we sent an online survey to a random sample of 27,000 physicians from 216,350 PCPs in the AMA database. Using case scenarios for atherosclerotic cardiovascular disease (ASCVD) and prostate cancer (PrCa) prevention, the survey asked how respondents would manage patients with no PRS (usual care) or a PRS indicating high or low risk. Respondents randomly received 1 of 4 surveys, which varied only by patient race in the ASCVD and PrCa cases [European- (white) or African-American (Black)]. Repeated measures across scenarios were analyzed using generalized estimating equations.
Results: Of 363 respondents (response rate 1.3%), mean (SD) age was 55 (13) years, 40% were women, and 35% reported non-white race/ethnicity. For ASCVD, PCPs were 5.2 times (95% CI 3.5-7.7) more likely and 49% (95% CI 36%-59%) less likely to prescribe a statin to patients with a high-risk and low-risk PRS, respectively, compared to usual care (both p0.05). Similarly, for PrCa, PCPs were 18 times (95% CI 7.5-42.8) more likely and 41% (95% CI 24%-55%) less likely to recommend prostate-specific antigen (PSA) screening for patients with a high-risk and low-risk PRS, respectively, compared to usual care (both p0.05). A majority identified cost (84%), insurance discrimination (64%), and lack of clinical guidelines (63%) as severe or moderate barriers to the use of PRS; only 34% identified PCP time as a barrier. Less than half (40%) somewhat or strongly agreed they felt confident in their clinical use of PRS.
Conclusions: A national sample of PCPs reported that PRS results would change their decision-making around disease prevention. We found no evidence that PCPs considered the variable performance of PRS by ancestry in using PRS results to change medical decision-making. Health system-level barriers may impede PRS uptake.

PrgmNr 2974 - Population carrier frequency estimates may improve genetic counselling for rare autosomal recessive syndromes due to biallelic variants in cancer risk genes

View session detail

Author Block: **J. Powers**[1], E. Trujillo[1], L. Conway[1], K. N. Maxwell[2]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ Pennsylvania, Philadelphia, PA
Disclosure Block: **J. Powers:** None.
Background: Several genes lead to different disease states depending on pattern of inheritance. Germline pathogenic variants (PV) in a subset of cancer predisposition genes not only associate with adult-onset autosomal dominantly (AD) inherited cancer risk in the monoallelic state, they also lead to life-limiting autosomal recessive (AR) childhood diseases when present in the biallelic state. However, prior studies have suggested that genetic counselling in cancer risk clinics often does not include the reproductive risk implications associated with the carrier state of a PV in cancer risk genes (CRG).**Methods:** To better facilitate genetic counselling regarding the reproductive risk implications associated with AD inherited CRGs, we used population-based data to calculate the estimated carrier frequencies in different ethnicities for genes that lead to AD cancer risk/AR disease. Genes were selected if they had: 1) well-established AD cancer risk, and 2) known causative role in AR disease. All variants in selected genes were identified in gnomADv.3.0 and classified according to ACMG rules. Population carrier frequencies were collapsed by gene and determined by summation of allele frequencies assuming Hardy-Weinberg equilibrium. Language found on reports from five clinical diagnostic laboratories was reviewed.**Results:** Carrier frequencies in seven ancestry cohorts were determined for *ATM, BRCA1, BRCA2, BRIP1, FH, MLH1, MSH2, MSH6, NBN, PMS2, RAD51C,* and *RAD51D*. *BRCA1/2* in the Ashkenazi Jewish (AJ) population had the highest carrier frequency at 1.32% and 1.29%, respectively. Excluding AJ, carrier frequencies across ethnicities were 0.19±0.08% (range 0.06-0.28%) for *BRCA1* and 0.31±0.10% (range 0.17-0.43%) for *BRCA2*. *ATM* had the next highest average carrier frequency at 0.31±0.12% (range 0.16-0.48%). *MLH1* and *MSH2* had the lowest, both at 0.03±0.01% (range 0.01-0.04% for both genes). Average carrier frequencies were less than 0.10% for *FH* and *RAD51D*, and between 0.10-0.20% for the remaining. On review of test reports, all labs described reproductive implications of the AR disease associated with the CRG. No labs provided discrete carrier frequency data nor disease prevalence. Oonly one lab discussed possible need for carrier testing. **Conclusions:** More robust characterization of carrier frequencies will better assist genetic counselors in facilitating discussions about reproductive risks of AR diseases in the setting of a PV in any above-mentioned CRGs, assuming their partner is not a known carrier. Future studies are needed to assess genetic providersâ behaviors surrounding discussion of risks as well as patient comprehension.

PrgmNr 2975 - Preparing primary care providers to utilize direct-to-consumer genetic test results in the creation of an individualized healthcare plan

View session detail

Author Block: **K. Glenn**; Clemson Univ./VCOM, Clemson, SC
Disclosure Block: **K. Glenn:** None.
Direct-to-consumer genetic testing (DTCGT) companies offer the average person a chance to unlock their genome to explore their ancestry, genetic traits, and health risk reports. The results always come with a stern suggestion to contact a healthcare provider if there are any questions or concerns. However, most primary care providers do not have a genetics-focused practice and may find themselves unsure of what to do with the unsolicited test results of a concerned patient. The purpose of this study is to explore medical studentsâ knowledge retention of DTCGT and how it uses genetic risk assessment in the formulation of test results. Using a pre and post-module assessment, this study aims to determine the effectiveness of an online course that trains future providers to use patient-initiated direct-to-consumer genetic test results in creating a personalized healthcare plan.

PrgmNr 2976 - Prevalence and Correlates of Parental Decisions about Direct-to-Consumer Genetic Testing for Adult-Onset Inherited Cancer Syndromes in their Children

View session detail

Author Block: **G. McDonnell**[1], B. N. Peshkin[1], J. G. Hamilton[2], J. Brower[3], H. Ovadia[2], L. Friedman Ross[4], R. Sacca[2], B. A. Tarini[5], B. S. Wilfond[6], S. M. Domchek[3], S. Vittone[7], S. Campbell[1], K. P. Tercyak[1]; [1]Georgetown Lombardi Comprehensive Cancer Ctr., Washington, DC, [2]Mem. Sloan Kettering Cancer Ctr., New York, NY, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]Univ. of Chicago, Chicago, IL, [5]Children's Natl. Hosp., Washington, DC, [6]Seattle Children's Hosp., Seattle, WA, [7]Georgetown Univ. Sch. of Nursing and Hlth.Studies, Washington, DC

Disclosure Block: **G. McDonnell:** None.

Direct-to-Consumer (DTC) genetic testing for mutations in adult-onset inherited cancer syndromes (e.g., breast, ovarian, colorectal cancers) is available. Although parentsâ□□ use of DTC testing to identify their childrenâ□□s risk status is controversial, it is precluded neither by some companiesâ□□ practices nor federal regulations. At present, very little is known about parentsâ□□ decisions on DTC testing for their children. This study examined pediatric DTC testing attitudes and behaviors in parents and correlates of testing in their children. The sample was drawn from a research registry, with invitations sent to N=591 adults with a pathogenic variant in a target gene (BRCA1/2, MLH1, MSH2/6, PMS2, MUTYH): N=142 likely had a child in the study age range. Overall, N=134 adults visited the study website, and 80 parents (60% of visitors; 95% female, 94% white) with children 10-21 y.o. participated. The survey included validated measures of pediatric genetic testing knowledge, attitudes, and beliefs, general and cancer-specific parent-child communication, and opinions about the pros/cons of DTC pediatric testing. In parents, 83% informed their children of the parentâ□□s mutation, and 56% talked openly about pediatric testing. Informing children about parental risk and talking about testing were highly associated (R=0.52, p=65): cons were lack of healthcare provider engagement (Mdn=19), lack of required genetic counseling (Mdn=22), and unmet psychosocial needs (Mdn=38). In sum, high risk parents commonly inform children about heritable cancer risks, most discuss genetic testing, and 10% also test children by DTC. Genetics professionals may see an increase in children referred for service after DTC testing. Outcome studies are needed to better inform practice and policy in this area.

PrgmNr 2977 - Public knowledge, perceptions, and attitudes towards direct-to-consumer genetic testing: A national web-based survey

View session detail

Author Block: H. Andrighetti, J. Furnival, J. Hunsanger, S. Bradley, A. Baker, B. Knorr, **S. D. Crawford**; SoundRocket, Ann Arbor, MI
Disclosure Block: **S.D. Crawford:** None.
The next decade will hold vast advances in genomics, and with it, genetic testing will become more common, both clinically and direct-to-consumer (DTC). DTC genetic testing (DTCGT) has progressed from a focus on ancestry and traits (so-called â□□recreationalâ□□ testing) to a range of offerings that increasingly impact medical care, up to and including whole genome sequencing. As this testing becomes both more accessible and clinically relevant, it is important to study public knowledge, perceptions and attitudes of DTCGT, and how these shift over time.
In May 2021, we conducted a national study that included a random sample of over 750 participants selected to match the U.S. Census demographics on gender, age, race/ethnicity, and geographic region. In this self-administered web-based survey, participants were asked about their familiarity with genetic terminology, experiences with genetic testing, and opinions on DTCGT. To examine current perceptions of DTCGT, participants were asked how strongly they agree with a list of 13 potential barriers to DTCGT (e.g.: concerns about privacy of genetic information) and 10 potential motivators for DTCGT (e.g.: learning about risks for specific genetic conditions).
We will describe the current landscape of genetic testing experiences among a nationally representative sample, and explore the most commonly endorsed current barriers and motivators to pursuing DTCGT. Additional analyses will include discussion of changes in perspectives over time, and comparisons based on age, gender, race/ethnicity, marital status, education level, proactive health behaviors, family history, genetics knowledge, and genetic testing experiences.
These data will serve as a snapshot of the current landscape of public knowledge, perceptions, and attitudes towards DTCGT, and as a baseline for an ongoing cross-sectional study to monitor how public perspectives change over time as the DTCGT space evolves. This exploratory study may give insights to those working both in the DTCGT space, and in clinical genetics, identifying knowledge gaps that may exist and ideal targets for public education and outreach as these tests become increasingly ubiquitous. In sharing these initial results, we hope to initiate a conversation around the appropriate priorities for our next stage of research.

PrgmNr 2978 - Returning Polygenic Risk Score-Based Results to Parents - Parentsâ Understanding of and Response to Reported Risk

View session detail

Author Block: **J. J. Connolly**[1], I. A. Holm[2], M. Behr[1], M. del Rosario[3], H. Hain[1], M. H. Harr[4], H. Hakonarson[4]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Boston Children's Hosp, Boston, MA, [3]Boston Children's Hosp., Boston, MA, [4]Children S Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **J.J. Connolly:** None.
Polygenic risk scores (PRSs) have the potential to improve healthcare by identifying individuals at risk for common complex conditions. Use of PRSs in clinical practice, however, requires careful assessment of the needs and capabilities of patients, providers, and healthcare systems. To this end, the electronic Medical Records and Genomics (eMERGE) network is conducting a pragmatic trial, which will return PRSs to 25,000 pediatric and adult participants. All participants will receive a risk report that includes a PRS calculation classifying them as either high risk (~3% per condition) or not. Returning PRSs to children is particularly novel, and eMERGE is leading a number of studies to inform best practices for PRS-based reporting in pediatrics. Here, we performed qualitative analyses from semi-structured interviews with 48 parents who received hypothetical risk reports, wherein children were classified as high risk for type 2 diabetes and asthma. Parents were assessed on their comprehension of absolute- versus relative-risk framing (overall risk perception); 2) likelihood to follow risk reduction recommendations; 3) perceived value of the information presented; 4) psychosocial impact (stress/anxiety); and 5) education/support needed. Analyses revealed miscomprehension of several important study parameters across all ethnic groups, and highlight the need for targeted educational material to formatively support report development.

PrgmNr 2979 - Secondary Findings in Minors Case Series: Recommendations for Clinical Care and Research

View session detail

Author Block: **M. Similuk**[1], A. Katz[2], M. R. Setzer[1], S. Miner[3], J. Yan[1], M. Kamen[4], C. Jodarski[4], L. Jamal[3,5], K. Jevtich[1], Y. Yu[1], R. Duncan[1], B. Seifert[4], R. Ghosh[4], S. Holland[1], M. A. Walkiewicz[1]; [1]Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [2]Natl. Human Genome Res. Inst., Bethesda, MD, [3]NIH Clinical Ctr. Dept. of Bioethics, Bethesda, MD, [4]Med. Sci. & Computing at the Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [5]Natl. Cancer Inst., NIH, Bethesda, MD
Disclosure Block: **M. Similuk:** None.
Background and Purpose: Increasingly, secondary findings (SF) are being returned to individuals in clinical and research settings. As an opportunistic screening result, a SF infers a lower positive predictive value for disease compared to a sequencing result identified through diagnostic testing, and as such clinicians can benefit from structured guidance on navigating the complexities of SF management. While Katz and colleagues recently articulated a SF management framework for adults (PMID: 32619490), there is little, and often conflicting, professional guidance for how SFs should be managed in pediatric settings. There is a need for clearer guidance given special nuances including the inability for the proband to provide consent, the possible value that this information may have for the proband and other family members, and developmental considerations for both the medical and psychosocial case management. In the National Institutes of Allergy and Infectious Diseases (NIAID) Centralized Sequencing Program at the National Institutes of Health Clinical Center, we have elected to return SF from exome and genome sequencing in minors by default. Here we synthesize specific findings from our cohort with empirical evidence and guidance from the literature.
Methods: 864 pediatric participants received exome or genome sequencing with SF analysis from 2017-2021.
Results: Seventeen SF were returned in 16 minor patients (1.9%). Three-quarters of the SF (13/17, 76.5%) were associated with potential childhood onset of disease. SFs in the following risk categories were identified: disorders with high or near complete penetrance (*FBN1*, *RET*; 4/17, 23.5%), disorders with low or moderate penetrance (*BRCA2*, *DSC2*, *MHY7*, *PKP2*; 8/17, 47.1%), disorders with low intensity interventions (*LDLR*, 4/17, 23.5%), and disorders with typical childhood presentations (*OTC*; 1/17, 5.9%).
Discussion: We present three primary themes with illustrative cases to demonstrate the nuances encountered in pediatric SF: (1.) Medical evaluation and risk assessment: SF interactions with indication for testing (Case A: Primary finding - *CYBB;* SF - *OTC*). (2.) Considerations for disclosures: family understanding and adaptation to SF (Case B: Primary finding - *CTLA4;* SF - *BRCA2* and *LDLR*) (3.) Opportunity for adult relatives at risk (Case C: SF - *PKP2*).
Conclusion: We synthesize guidance on adult SF management, the evolving empirical literature on pediatric adjustment, and genetic testing ethics to create an illuminating pediatric SF case series with recommendations for clinical care and future research.

PrgmNr 2980 - Strategies to address the health and social care needs of rare disease patients: A comparative analysis of two Canadian provinces

View session detail

Author Block: **W. Lee**, R. Z. Hayeems; The Hosp. for Sick Children, Toronto, ON, Canada
Disclosure Block: **W. Lee:** None.
Background: Rare diseases are a serious public health challenge that affects not only patients and families, but also the health care system and society. A comprehensive approach to rare diseases is needed to address gaps that exist in diagnosis, prevention, care, treatment, social support, and research. While many countries have developed national policies and strategies, Canada is lacking a comprehensive national strategy for addressing the unmet health and social care needs for these patients. In 2015, a national rare disease advocacy group, Canadian Organization for Rare Disorders (CORD), introduced Canadaâ□□s Rare Disease Strategy. Their final report identified five areas as high priority in rare disease, but CORDâ□□s recommendations have not been fully endorsed or adopted by the federal government. While a national strategy is lacking, several Canadian jurisdictions have developed and implemented their own policies and programs that address one or more of CORDâ□□s five key areas.
Methods: Using the World Health Organization Health Systems Framework, which outlines the five essential functions of health systems, we conducted a comprehensive review of the academic and grey literature to describe and compare the approaches to rare disease care in two provinces: Alberta and Ontario. Using directed content analysis, CORDâ□□s five priority areas were used to describe the extent to which existing provincial programs address the gaps and challenges faced by the rare disease community.
Results: Documents included in the comparative analysis included government reports, policy documents, reports from professional associations or patient advocacy groups, peer-reviewed articles, and media reports available as of April 5, 2021. The comparative analysis revealed that while both provinces have similar programs in place in terms of their objectives, there is evidence of more comprehensive patient access to diagnostic services and essential medicine in Ontario compared to Alberta. While understanding the reasons underlying the discrepancies between the provinces was not one of the main objectives of this analysis, differences in the health care system structure (i.e., centralized vs. decentralized) and the number of rare disease patients in the two provinces may be contributing factors. The analysis also identified common challenges across the two provinces related to workforce shortages and limited health information systems.
Conclusion: The learnings from this comparative analysis combined with international rare disease strategies can inform the development of a pan-Canadian rare disease strategy.

PrgmNr 2981 - The complexity of diagnosing rare disease: An organising framework for health services and health economics research based on real world evidence

View session detail

Author Block: **R. Z. Hayeems**[1,2], C. Michaels-Igbokwe[3], V. Venkataramanan[1], T. Hartley[4], M. Acker[1], M. Gillespie[4], W. J. Ungar[5,2], R. Mendoza-Londono[5,6], F. P. Bernier[7,8], K. M. Boycott[9,10], D. A. Marshall[3,11]; [1]Child Hlth.Evaluative Sci., The Hosp. for Sick Children, Toronto, ON, Canada, [2]Inst. of Hlth.Policy, Management and Evaluation, The Univ. of Toronto, Toronto, ON, Canada, [3]Cumming Sch. of Med., Univ. of Calgary, Calgary, AB, Canada, [4]Dept. of Genetics, Childrenâ□□s Hosp. of Eastern Ontario, Ottawa, ON, Canada, [5]The Hosp. for Sick Children, Toronto, ON, Canada, [6]Dept. of Pediatrics, The Univ. of Toronto, Toronto, ON, Canada, [7]Alberta Childrenâ□□s Hosp., Calgary, AB, Canada, [8]Alberta Childrenâ□□s Hosp. Res. Inst., Univ. of Calgary, Calgary, AB, Canada, [9]Childrenâ□□s Hosp. of Eastern Ontario, Ottawa, ON, Canada, [10]Childrenâ□□s Hosp. Eastern Ontario Res. Inst., Univ. of Ottawa, Ottawa, ON, Canada, [11]Oâ□□Brien Inst. for Publ. Hlth., Univ. of Calgary, Calgary, AB, Canada
Disclosure Block: **R.Z. Hayeems:** None.
Purpose: To facilitate robust health services and economic analyses of clinical exome and genome sequencing, the purpose of this study was to establish a framework for organizing diagnostic testing trajectories for rare disease patients. **Methods:** Embedded within a large Canadian rare disease cohort study (Care4Rare SOLVE), we collected diagnostic investigations prior to exome sequencing from medical records of 228 patients. Medical geneticist experts participated in a five-step consensus-building process to develop a real-world data-driven framework for organizing the complex range of observed tests. Experts categorized tests as *indicator* or *non-indicator* based on their specific contribution to diagnosing rare disease. In addition, indicator tests included those that were higher cost, potentially invasive, less accessible, and ordered/interpreted by a sub-specialist and non-indicator tests included those that were lower cost, non-invasive, locally accessible, and ordered/interpreted by a generalist. Face validity of the resulting SOLVE Framework was assessed using case vignettes that spanned a range of phenotypic indications for testing. **Results:** Most cases had symptom onset at birth (42.5%) or during childhood (43.4%) and had intellectual disability (73.3%). On average, the time spent seeking a diagnosis prior to sequencing was 2290 days [SD=2138] and included up to 71 diagnostic tests. Agreement across experts on indicator vs. non-indicator test categories ranged from 83%-96%. Observed tests, including 186 unique indicator and 39 unique non-indicator tests across cytogenetic/molecular (n=112), biochemical (n=73), imaging (n=28), electrical (n=7), and pathology (n=5) test categories comprised the SOLVE Framework. **Conclusion:** Real world diagnostic testing data can be ascertained and organized to reflect the complexity of the journey for patients with rare disease. Informed by a robust sample of 228 rare disease patients from multiple clinical settings, the SOLVE Framework will improve the accuracy and certainty associated with economic assessments of genomic sequencing by providing a consistent approach to categorising and comparing patient trajectories. The application of this framework to the economic evaluation of the complete SOLVE cohort is underway.

PrgmNr 2982 - The cost of good health: Poverty association with differential gene expression

View session detail

Author Block: **N. Arnold**[1,2], J. Resztak[2], A. Alazizi[2], S. Dubaisi[2], R. J. Thorpe Jr[3], N. Noren Hooten[4], M. K. Evans[4], D. F. Dluzen[5,1], R. Pique-Regi[6], F. Luca[6]; [1]Morgan State Univ., Baltimore, MD, [2]Wayne State Univ., Detroit, MI, [3]John Hopkins Univ., Baltimore, MD, [4]Natl. Inst. on Aging, NIH, Baltimore, MD, [5]NIH, Silver Spring, MD, [6]Wayne State Univ, Detroit, MI
Disclosure Block: **N. Arnold:** None.
PURPOSE Psychosocial factors exert a powerful influence on health status and contribute to health disparities among marginalized populations. For example, overall life expectancy at birth throughout the United States tracks with poverty level, educational attainment, economic security and other upstream social determinants of health. Socioeconomic status (SES) and psychosocial factors are documented to affect gene expression in peripheral blood mononuclear cells, suggesting a molecular mechanism for some health disparities. Here we investigated the effects of poverty among Baltimore City residents participating in the Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS). **METHODS** We examined 239 participants of the HANDLS cohort study in Baltimore whose reported household income was either above or below the 125% federal poverty line for 2004. This population sample was composed of 119 African Americans and 120 white, for a total of 119 men and 120 women. We performed RNA sequencing in peripheral blood mononuclear cells to assess differential gene expression patterns associated with poverty. **RESULTS** We identified 15 genes differentially expressed when testing for poverty while controlling for race, sex, and age. When focusing on women, we found that individuals living in poverty had increased expression for 9 genes and decreased expression for 11 genes compared to individuals living above the poverty line. GSEA identified an enrichment for Herpes simplex virus infection pathway and B cell mediated immunity in genes differentially expressed in women living in poverty. **CONCLUSIONS** Our study suggests that poverty status influences gene expression in the immune system. Improving health outcomes for at-risk populations is achievable by understanding the link between poverty and identifiable biological mechanisms that influence disease.

PrgmNr 2983 - The Effect of Genetic Testing as Part of Personalized Lifestyle and Habit Change Coaching

View session detail

Author Block: **L. Fensterheim**[1], N. Vani[1], J. Ruby[1], K. Hood[1], L. Dodo[1], R. Bennie[1], L. PÃ©russe[2]; [1]Newtopia Inc., Toronto, ON, Canada, [2]Laval Univ., Quebec, QC, Canada
Disclosure Block: **L. Fensterheim:** Salary/Employment; Newtopia Inc.
Background Genetic testing can be used as part of lifestyle intervention programs to provide personalized recommendations with the hope of improving adherence and facilitate changes in lifestyle behaviors. However, little is known about the impact of such testing on the effectiveness of a lifestyle coaching experience. **Objective** To evaluate whether targeted genetic testing improves weight loss outcomes from a personalized lifestyle and habit change coaching experience. **Methods** This was a retrospective study of participants who engaged for at least 4 months in a personalized habit change coaching program between January 1, and February 28, 2020. Participants were required to have an initial body mass index (BMI) of at least 28 kg/m^2. All participants were offered genetic testing to optimize the delivery of the personalized coaching. The following three candidate gene variants were assessed based on their documented association with obesity: DRD2/ANKK1 rs1800497, FTO rs9939609 and MC4R rs17782313. Two main outcomes were considered: percent weight loss achieved at 12 months (%WL12) and achieving a minimum of 5% weight loss at 12 months (5%WL12). A generalized linear model was used to assess whether %WL12 differed between those who received genetic testing versus those who did not. A logistic model was also used to assess the impact of genetic testing on 5%WL12. Age, sex, ethnic background, initial weight, and personality type were controlled for in all analyses. **Results** A total of 2297 participants (52% female) met the criteria for inclusion. On average (mean Â± SD), participants were 45.4 Â± 10.7 years and had a BMI of 34.1 Â± 5.7 kg/m^2, respectively. The covariate adjusted %WL12 was significantly higher (p = 0.001) in those taking the test (4.0 Â± 0.3) compared to those who did not (3.2 Â± 0.2) The likelihood of achieving 5%WL12 had 1.3 times higher odds in participants taking the genetic test (OR= 1.3; 95%CI: 1.08-1.59) versus those who did not. **Conclusion:** These results suggest that targeted genetic testing can improve the effectiveness of a lifestyle coaching program when properly integrated in terms of the weight loss achieved in response to the experience.

PrgmNr 2984 - The physical costs and psychosocial benefits associated with moving to a White neighborhood for African American male children

View session detail

Author Block: **J. Del Toro**[1], L. Gaydosh[2], M-T. Wang[1], L. Schneper[3], C. Mitchell[4], S. McLanahan[3], D. A. Notterman[3]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Vanderbilt Univ., Nashville, TN, [3]Princeton Univ., Princeton, NJ, [4]Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **J. Del Toro:** None.
For youth of color, moving to an affluent neighborhood introduces a paradox: Positive psychosocial adjustment in exchange for physiological dysfunction. According to skin-deep resilience (Brody et al., 2013), persistence in achieving upward social mobility produce favorable psychosocial outcomes yet also predispose youth of color to racial discrimination, which over time increases allostatic load (Chen et al., 2015), metabolic syndrome (Miller et al., 2020), and likelihood of developing diabetes (Brody et al., 2016). However, studies examining risk exposure for such youth have relied on colorblind approaches as they have not acknowledged that upward social mobility entails shifting into a predominantly White context. Moreover, socioeconomic inequality is problematic for African American male children as those who grow up economically advantaged are more likely to fall to the bottom of the socioeconomic ladder and those at the bottom of the socioeconomic latter are less likely to climb up relative to other racial and gender groups (Badger et al., 2018). Thus, to extend the literature, we examined the physiological and psychosocial consequences associated with moving to a White neighborhood for African American and White male children in the US.
950 African American and 331 White males participated in two waves of the Fragile Families and Child Wellbeing Study, a national cohort study of children primarily from unwedded parents. Children participated when they were on average ages 9 and 15. Data at each wave included the U.S. Census tract percentage of White residents, self-reported depressive symptoms and delinquency, and salivary telomere length. Using our data, we estimated multi-level models in which we nested time within participants and accounted for city random effects.
After controlling for covariates, increases in the percentage of White residents predicted decreases in depressive symptoms and delinquency between ages 9 and 15 among African American males. However, increases in the percentage of White residents predicted telomere length shortening between ages 9 and 15 for African American males. Changes in the percentage of White residents did not predict depressive symptoms, delinquency, or telomere length for White males.
We found that the racial composition of the neighborhood context is crucial in shaping youthâ□□s developmental trajectories. While we support propositions that upward social mobility can predispose youth of color to racial discrimination, acknowledging the racial composition of the context is necessary to understand the dynamics that youth face when their families aim to live a better life for their children.

PrgmNr 2985 - The White Ceiling: Qualitative Study of Genetic Counseling Studentsâ Perceptions about Cultural Competence Training

View session detail

Author Block: **T. Rai**, S. Capasso, J. Duffy; Bay Path Univ., Longmeadow, MA
Disclosure Block: **T. Rai:** None.
Objective: This qualitative research was conducted to study the perception of current and recently graduated genetic counseling students from the classes of 2018 to 2022, on the CCT provided by their programs. According to the NSGCâs own Professional Status Survey, over the last 20 years, the population of professional genetic counselors who identify as White/Caucasian has held steady at 90% or more, despite years of JEDI focused efforts by the NSGC, other professional groups, and individuals. It is evident that it will be a while before we see significant results. The next best tool in ensuring patient centered care is delivered to our ever-diversifying patient population is by providing quality and effective cultural competence training (CCT) to our students.
Methods: A survey comprising 15 questions directed at answering the research question of the studentsâ perceptions of the type, quality, and effectiveness of CCT was created. Invitation for participation was sent through the NSGC listserv, social media, program directors, professional organizations, and word of mouth. After a 6 weeks collection period, 73 valid responses were received representing 33 of the 55 total US and Canadian Accreditation Council for Genetic Counseling (ACGC) accredited schools. The responses were uploaded into Dedoose ver 8.3.47b, a qualitative analysis software, and the material was scoured for common themes.
Results: The participants were from the classes of 2018 (*n*=3, 4.1%), 2019 (*n*=7, 9.6%), 2020 (*n*=13, 17.8%), 2021 (*n*=34, 46.6%), and 2022 (*n*=16, 21.9%), and 78% of the participants had more than a year of shadowing and direct patient counseling experience. Pre-CCT, students professed to having political, religious, SES, cultural/ethnicity/race, gender, and disability biases. The sub-themes that emerged from this study focused around the overarching themes of diversity and self-awareness. Students highlighted the lack of diversity in course topics and materials, the representative populations that are highlighted, invited speakers, and lecturers and program leadership. Students also expressed how programs failed to be self-aware by using language that fell into the trap of stereotyping and generalizing, lecturers failing to address and provide spaces to discuss xenophobic, racist, and culturally insensitive comments made in class, failing to provide opportunity to discuss course material, not making CCT classes or events a priority, and by overcompensating and causing students to dread going into clinic for fear of saying the wrong things.

PrgmNr 2986 - Views on the training needs of genetic assistants vary widely

View session detail

Author Block: **R. Rider**[1], L. Cubano[1], E. B. Madden[2], K. Jacoby Morris[3], R. Rowley[4], T. A. Manolio[2]; [1]Div. of Genomic Med., NHGRI, Bethesda, MD, [2]NHGRI, Bethesda, MD, [3]Ed. and Community Involvement Branch, NHGRI, Bethesda, MD, [4]Div. of Genomic Med., Bethesda, MD
Disclosure Block: **R. Rider:** None.
Introduction: Genetic assistants (GAs), often called â□□genetic counseling assistants,â□□ are a growing part of the workforce. We sought to identify opinions on the training needs of GAs and attitudes about GA training programs (GATPs). **Methods:** Links to a survey on training needs of GAs and attitudes about GATPs were emailed to members of the National Society of Genetic Counselors, 16 state genetic counseling (GC) professional organizations, and genomic medicine researchers. Proportions of responses with 95% confidence intervals were calculated and compared. Results were stratified by profession (GA vs. GC) and work setting (clinical vs. lab). Open-ended responses were coded to identify emerging themes. **Results:** Respondents (*n*=375) primarily included GCs (*n*=211) and GAs (*n*=125). While 73% had at least one GA in their work setting with a bachelorâ□□s degree, 16% worked with at least one GA with an associate degree or lower. For those in work settings that require GAs to have a bachelorâ□□s, similar numbers favored keeping that requirement vs. hiring someone without a bachelorâ□□s if they attended a GATP (53% [45-60] vs. 47% [39-55]). Although 37% preferred hiring those who completed GATPs, 55% were neutral. Most believed GATPs should focus on general background knowledge, with almost all skills to be learned on the job. Most skills (65%) were rated as unneeded by 25%-45% of respondents. While 47% felt that GATPs should require enrollment in or completion of a bachelorâ□□s, 29% felt an associate or lower would be appropriate and 11% felt that programs were needed at multiple levels. Almost half (44%) believed GATPs could be a burden that discourages candidates from becoming GAs. Minor differences by profession and work setting emerged, *e.g.*, those in clinical settings tended to rate GATPs as helpful more often than those in labs (67% [61-72] vs. 52% [40-63]). In the qualitative data, the most common theme was the concern that GATPs could become barriers to becoming a GA or a GC. Some worried that these barriers could ultimately reduce workforce diversity in genetic counseling. The next most common themes were that GAs in different settings have different training needs and that GATPs may be unnecessary, *e.g.*, repetitious of prior education or superfluous. **Conclusions:** While many agreed that GATPs could be useful, there was no consensus on specifics such as general knowledge and skills needed by GAs, target GATP audience, or ideal GATP content. Neither profession nor work setting explained the range of opinions about training needs or the concern that GATPs could become barriers. Better understanding of the nuances of GA training needs is vital to designing GATPs.

PrgmNr 2987 - Yield of genomic disease given the identification of a secondary finding

View session detail

Author Block: **A. Katz**[1], H. Shiferaw[2], C. Hong[3], L. G. Biesecker[4]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]NIH, Gaithersburg, MD, [3]Natl. Inst.S OF HEALTH, BETHESDA, MD, [4]Bethesda, MD
Disclosure Block: **A. Katz:** None.
A secondary finding is a genomic variant of potential medical value that is unrelated to the primary reason for testing. Here we define yield as the fraction of those with confirmed disease (clinicomolecular diagnosis) divided by those with a secondary finding. The yield of a secondary finding depends on the probability of pathogenicity of the variant (used here as a conditional probability) and the prior probability of the disorder. Using a Bayesian framework, we determined the posterior probability of disease (yield) given the identification of a secondary finding for each disorder associated with the ACMG recommended list of secondary findings genes. The hypothetical scenario we evaluated is an individual (for whom no clinical information is available) has received a secondary finding result. To determine the posterior probability of disease based on a genomic result alone (yield), we estimated the following parameters for each disorder: 1) the prior probability, which is the population prevalence of the disorder; 2) the probability that a pathogenic variant is identified by diagnostic sequencing in an individual who meets clinical criteria for the disorder; 3) the probability that an unaffected individual harbors a pathogenic variant associated with the disorder. Best estimates for parameters 1) and 2) were determined via literature review. Parameter 3) was determined by the frequency of individuals whose sequence data are in gnomAD who harbor variants that meet ACMG criteria for Pathogenic or Likely Pathogenic (with adjustments for disorders in which affected individuals may be included in gnomAD, such as hereditary breast and ovarian cancer syndrome). We focused on disorders displaying an autosomal dominant inheritance pattern. Applying Bayes theorem with the parameters noted, we calculated the posterior probability of disease when a pathogenic variant is identified in a genomic screening context. Among the ACMG recommended list, variants in familial hypercholesterolemia (prob=60%, RR=120.4 [95% 64.7-223.9]) had the highest yield and hereditary paraganglioma-pheochromocytoma syndrome (prob=0.065%, RR=400.2 [95% 25.0-6395.2]) had the lowest yield. As expected, the yield varies greatly depending on the population prevalence of the disorder. Our results are consistent with the notion that individuals with secondary findings have a high relative risk of disease (compared to the general population) but also underscore the distinction of a secondary finding from a genomic disease. Our calculated posterior probability of disease (yield) can be useful in the evaluation and management of individuals with secondary findings.

PrgmNr 2988 - A 3-part Phase 2 study of HST5040, an investigational oral therapy that reduces toxic coenzyme A metabolites in methylmalonic and propionic acidemias (clinicaltrials.gov NCT04732429)

View session detail

Author Block: **G. F. Cox**[1], M. Y. Waller[1], A. J. Armstrong[1], M. P. Hayes[1], K. A. Chapman[2]; [1]HemoShear Therapeutics, Charlottesville, VA, [2]Children's Natl. Hosp., Washington, DC
Disclosure Block: **G.F. Cox:** Major Stockholder/Ownership Interest; CANbridge, Deep Genomics, HemoShear. Consultant/Consulting Fees/Other Remuneration; Blueprint, CANbridge, Chiesi, Exonics/Vertex, Sanofi Genzyme, Ultragenyx.
Methylmalonic and propionic acidemias (MMA and PA) are organic acidemias caused by sequential enzyme deficiencies in the catabolism of propionic acid generated from certain amino acids (valine, methionine, isoleucine, and threonine), odd-chain fatty acids, cholesterol, and gut flora. Both disorders are characterized by frequent acute metabolic decompensations in early childhood involving metabolic acidosis and hyperammonemia, and long-term sequelae including intellectual disability, renal insufficiency (MMA>PA), and cardiac disease (arrhythmias and cardiomyopathy, PA>MMA). Current treatments do not address the underlying cause of these multisystemic diseases, although liver and/or kidney transplant have been shown to reduce the frequency of acute metabolic decompensations and improve quality of life. HST5040 (2,2-dimethylbutanoic acid) is a small molecule that diverts coenzyme A (CoA) from propionyl-CoA pathways via the formation of HST5040-CoA. In cultured MMA and PA patient primary hepatocytes, HST5040 reduces propionyl-CoA, methylmalonyl-CoA, and the derived toxic metabolites: 2-methylcitric acid, methylmalonic acid, and propionyl-carnitine. HST5040 is a liquid formulation that may be administered orally once daily and has high bioavailability. Nonclinical studies have shown broad tissue and organ biodistribution of HST5040, including to the CNS. A Phase 2 study was initiated in March 2021 and plans to enroll 12 US patients (6 MMA-mutase and 6 PA) age 2 and older in a 3-part sequential study design. Part A is a within-patient, dose-escalation study to identify the optimal dose of HST5040 based on reductions in disease-related biomarkers and safety. Part B is a randomized, double-blind, placebo-controlled crossover study involving two 3-month treatment periods to confirm changes in biomarkers and identify acute clinical responses. Part C is an open-label extension study to collect long-term safety and efficacy data, and in addition, may enroll additional patients who are post-transplant or have CblA or CblB deficiency. The primary endpoint for Part A and B is the change in plasma 2-methylcitric acid, a toxic metabolite derived from propionyl-CoA that inhibits the TCA cycle and impairs energy production. Secondary endpoints include other disease-related biomarkers, frequency of acute metabolic decompensations, measures of cognitive, behavioral, cardiac, and renal function, oral intake, and quality of life. The Phase 2 study design was presented at the 2021 ACMG meeting, and an update will be provided. The study is posted on clinicaltrials.gov (NCT04732429).

PrgmNr 2989 - Characterization of CRISPR gene editing in the brain of ataxia mouse models with targeted PCR-free Nanopore sequencing

View session detail

Author Block: **B. P. Simpson**[1,2], C. M. Yrigollen[1], P. T. Ranum[1], B. L. Davidson[1,2]; [1]CCMT, Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA
Disclosure Block: **B.P. Simpson:** None.

Gene therapies using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) are being developed to treat a myriad of previously intractable genetic disorders. Adeno-associated virus (AAV) vectors are often used to deliver the Cas nuclease and guide RNA (gRNA) to the affected tissues and cells. The potential consequences of CRISPR editing includes small indels, large deletions, inversions, translocations, and integrations of the AAV genomic sequence. We are developing AAV-CRISPR strategies to treat spinocerebellar ataxia type 2 (SCA2), an autosomal dominant neurodegenerative disease caused by a CAG trinucleotide repeat expansion in *ATXN2*. The successful translation of gene editing therapies into the clinic necessitates accurate methods to interrogate and predict editing events.

High AAV vector integration into endogenous DNA following CRISPR-mediated cleavage in the hippocampus and peripheral tissues has been previously reported using PCR-based assays. To eliminate the size and sequence specific bias of amplification by polymerase-based sequencing, and to closely examine AAV integration sequences, we used Oxford Nanopore Technology to sequence the genomic DNA of SCA2 transgenic mouse models treated with AAV-CRISPR. Both mouse models contain either 72 or 127 CAG repeats within the human *ATXN2* transgene, thereby further making PCR of the GC-rich repetitive sequence inefficient and unreliable for assessing editing.

Native DNA from brain regions of mice injected with AAV-CRISPR was used to generate long read sequencing libraries. The human *ATXN2* transgene was targeted for sequencing using Cas-mediated enrichment. Nanopore MinION sequencing runs resulted in long reads with a mean length of ~10 kb and target coverage as high as 3500x (~1% aligned reads) after alignment to the mouse genome and transgenic allele. Mice treated with dual-gRNAs flanking the CAG repeat sequence were shown to contain deletions of the expected 460 bp size at a low frequency of ~3%, suggesting the majority of loci were unedited or repaired as small indels. We detected AAV-CRISPR delivered cargo sequence integration at the target locus with a frequency of ~10-13% for a single gRNA site. Integrated sequences ranged from whole, 4.8 kb AAV genomes to partial fragments with slightly higher coverage of AAV inverted terminal repeat (ITR) sequence.

PCR-free target enrichment and sequencing strategies offer an unbiased assessment of native and edited genomic sequence, important for assessing CRISPR-mediated DNA editing. Our Nanopore sequencing results are important for development and safety considerations of CRISPR gene editing therapies for SCA2 and other disorders.

PrgmNr 2990 - CRISPR targeting of the dominant*Coch*p.A449T mutation in a patient-derived skin fibroblast cell line and in a new humanized mouse model

View session detail

Author Block: **H. Romi**[1], N. G. Robertson[2], S. Vijayakumar[3], C. Gurumurthy[4], B. Kleinstiver[5], R. Sherwood[6], M. Ivanchenko[7], K. Booth[7], D. Corey[7], C. C. Morton[8]; [1]Boston Children's Hosp., Boston, MA, [2]Brigham and Women's Hosp. & Harvard Med Sch., Boston, MA, [3]Creighton Univ., Omaha, NE, [4]Univ. of Nebraska Med. Ctr., Omaha, NE, [5]Massachussetts Gen. Hosp., Boston, MA, [6]Brigham and Women's Hosp., Boston, MA, [7]Harvard Med. Sch., Boston, MA, [8]Brigham & Women's Hosp, Boston, MA
Disclosure Block: **H. Romi:** None.
Dominant gain-of-function mutations in *COCH* are causative of the adult-onset progressive sensorineural hearing loss (HL) and vestibular disorder, DFNA9. Our efforts are focused on disruption of an autosomal dominant *COCH* mutation using CRISPR-Cas9 strategies, specifically targeting the mutant allele, while keeping the normal allele intact, which would enable adequate hearing function. Our efforts are envisioned to have a broader scope in establishing methods for gene therapy in other forms of autosomal dominant HL with gain-of-function or dominant-negative pathomechanisms, as these tend to have late-onset presentation providing a window of opportunity for therapeutic intervention. We studied a family segregating post-lingual progressive autosomal dominant non-syndromic hearing loss with the novel missense variant p.A449T in *COCH*. Pedigree analysis confirmed segregation of p.A449T in all affected family members. Using a skin biopsy from a carrier of the p.A449T variant, we generated a fibroblast cell line. Through co-transfection with a plasmid carrying SaCas9-KKH and *COCH* guide-RNA expression plasmid, we were successful in creating indel disruption only in the mutant allele with an editing efficiency of 43% of the mutant allele, and no disruption of the normal allele. These *in vitro* studies are poised to be continued in our current *in vivo* targeting experiments. We have also created a humanized mouse model for the p.A449T variant (*Coch*$^{A449T/+}$). This is the first *Coch* mouse model to be established with a variant in the 3â□□ region of the vWFA domain. DFNA9 individuals with this 3â□□ variant display an earlier onset of hearing loss, typically in the first to second decades of life, as compared to individuals with variants in the more upstream 5â□□ LCCL domain of cochlin. Therefore, by creating the A449T mouse model, we will be able to conduct our CRISPR intervention studies pre- and post-hearing loss at earlier time-points compared to our existing mouse model with a variant in the 5â□□ region of the gene. Auditory Brainstem Responses (ABRs) as well as Distortion Product Otoacoustic Emissions (DPOAE) analyses of 9-month-old mice have shown significantly elevated thresholds for both *Coch*$^{A449T/+}$ (heterozygous) and *Coch*$^{A449T/A449T}$ (homozygous) mice, as compared to their *Coch*$^{+/+}$ (wild-type) littermates, at all tested frequencies. Mice at even younger ages are currently being evaluated. The present goal of this aspect of the project is to perform round window micro-injection of CRISPR-Cas9 and guide RNA into these mice and to measure hearing in different experimental groups, who receive treatment vs. sham injections.

PrgmNr 2991 - Haplotyping SNPs for allele-specific gene editing of the mutant huntingtin allele using long-read sequencing

View session detail

Author Block: **L. Fang**[1], A. M. Monteys[1], A. DÃ¼rr[2], M. Keiser[1], C. Cheng[1], A. Harapanahalli[1], P. Gonzalez-Alegre[1,3], B. L. Davidson[4], K. Wang[5]; [1]Children's Hosp. of Philadelphia, PHILADELPHIA, PA, [2]Sorbonne UniversitÃ©, Paris, France, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]The Children's Hosp. of Philadelphia, Philadelphia, PA, [5]Childrenâ□□s Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **L. Fang:** None.

Huntington's disease (HD) is an autosomal dominant neurodegenerative disease caused by CAG repeat expansions in the huntingtin (HTT) gene. Although trials of disease-modifying treatments are now on the horizon, the clinical care is focused on symptom management. We previously reported allele-specific deletion of the mutant HTT by CRISPR/Cas9 in a mouse model and human cell lines. Allele selectivity is achieved by targeting heterozygous SNPs that create or eliminate a Protospacer Adjacent Motif (PAM). However, given the lack of knowledge on haplotype structure in HD populations, a comprehensive analysis of all potential targeting sites is lacking and the optimal personalized editing strategy for HD individuals is unknown. To address this, we developed a multiplexed targeted long-read sequencing approach to sequence a 10.4 kb genomic region flanking exon-1 of HTT and created necessary computational tools (AmpBinner and AmpRepeat) to de-multiplex the data, detect repeats, and phase the reads. We applied this approach to two independent HD cohorts (974 individuals from the US and France), detected SNPs, analyzed haplotypes and potential editing sites for various enzymes. In the haplotype structure analysis, we showed potential founder effects in unrelated HD individuals from different continents. Based on the haplotype analysis, 23% of HD individuals of European ancestry can be edited by targeting one SNP (rs2857935). Up to 56% HD individuals of European ancestry can be potentially edited by combinatorial targeting of multiple SNPs. Our results provide the first haplotype map of the region surrounding exon-1 of HTT in HD cohorts. Our workflow can be applied to other repeat expansion diseases to facilitate allele-specific gene editing.

PrgmNr 2992 - Human genetic diversity modifies therapeutic gene editing off-target potential

View session detail

Author Block: **L. Y. Lin**[1,2,3], S. Cancellieri[4], J. Zeng[1,2,3], F. Masillo[4], M. Nguyen[1,2,3], N. Bombieri[4], S. A. Maitland[5], M-F. Ciuculescu[1], V. Katta[6], S. Q. Tsai[6], M. Armant[1], S. A. Wolfe[5], R. Giugno[4], D. E. Bauer[1,2,3], L. Pinello[7,3]; [1]Boston Childrenâ□□s Hosp., Boston, MA, [2]Dana-Farber Cancer Inst., Boston, MA, [3]Harvard Med. Sch., Boston, MA, [4]Univ. of Verona, Verona, Italy, [5]Univ. of Massachusetts Med. Sch., Worcester, MA, [6]St. Jude Children's Res. Hosp., Memphis, TN, [7]Massachusetts Gen. Hosp., Boston, MA

Disclosure Block: **L.Y. Lin:** None.

CRISPR gene editing holds great promise to modify somatic genomes to ameliorate disease. In silico prediction of homologous sites coupled with biochemical evaluation of possible genomic off-targets may predict genotoxicity risk of individual gene editing reagents. However, standard computational and biochemical methods focus on reference genomes and do not consider the impact of genetic diversity on off-target potential. Here we developed a web application called CRISPRme that explicitly and efficiently integrates human genetic variant datasets with orthogonal genomic annotations to predict and prioritize off-target sites at scale. The method considers both single-nucleotide variants (SNVs) and indels, accounts for bona fide haplotypes, accepts spacer:protospacer mismatches and bulges, and is suitable for personal genome analyses. We tested the tool with a guide RNA (gRNA) targeting the *BCL11A* erythroid enhancer that has shown therapeutic promise in clinical trials for sickle cell disease (SCD) and Î²-thalassemia (Frangoul et al. *NEJM* 2021). We find that the top predicted off-target site is produced by a non-reference allele common in African-ancestry populations (rs114518452, minor allele frequency (MAF) = 4.5%) that introduces a protospacer adjacent motif (PAM) for SpCas9. We validate that SpCas9 generates indels (~9.6% frequency) and chr2 pericentric inversions in a strictly allele-specific manner in edited CD34+ hematopoietic stem/progenitor cells (HSPCs), although a high-fidelity Cas9 variant mitigates this off-target. Our work illustrates how genetic variation may modify the genomic outcomes of therapeutic gene editing and provides a simple tool for comprehensive off-target assessment. CRISPRme is available at http://crisprme.di.univr.it.

PrgmNr 2993 - Long-term, sustained efficacy and safety from a phase 1/2 clinical trial of an AAV8-mediated liver-directed gene therapy in adults with glycogen storage disease type Ia

View session detail

Author Block: **J. Mitchell**[1], R. Riba-Wolman[2], A. Ahmad[3], M. L. Couce Pico[4], T. G. Derks[5], D. A. Weinstein[2], D. F. Rodriguez-Buritica[6], C. Lee[7], V. Valayannopoulos[7], E. Crombez[7]; [1]Montreal Childrenâ□□s Hosp., Montreal, QC, Canada, [2]Univ. of Connecticut, Farmington, CT, [3]Univ. of Michigan, Ann Arbor, MI, [4]Hosp. ClÃnico Univ.rio de Santiago de Compostela, Santiago de Compostela, Spain, [5]Univ. of Groningen, Groningen, Netherlands, [6]Univ. of Texas McGovern Med. Sch., Houston, TX, [7]Ultragenyx Pharmaceutical Inc., Cambridge, MA

Disclosure Block: **J. Mitchell:** Grant/Contracted Research Support (External); Ultragenyx Pharmaceutical Inc..

Glycogen storage disease type Ia (GSDIa) results from a deficiency of glucose 6-phosphatase (G6Pase) which is essential for glycogenolysis and gluconeogenesis. DTX401 is an adeno-associated virus serotype 8 (AAV8) vector expressing the human G6Pase gene (*G6PC*). This global, open-label, phase 1/2, dose escalation gene therapy trial (NCT03517085) is evaluating the safety and efficacy of a single DTX401 intravenous infusion in adults with GSDIa. Three patients in Cohort 1 received DTX401 2.0 x 10^{12} gene copies (GC)/kg, and three each in Cohorts 2, 3, and 4 received DTX401 6.0 x 10^{12} GC/kg. Cohort 4 was recently enrolled and includes a prophylactic steroid regimen to prevent transaminase elevation related to the immune response to DTX401. In the nine patients enrolled in Cohorts 1 through 3, mean (SD) total daily cornstarch intake reduction from baseline to Week 52 was 65.5% (22.3) and from baseline to last visit (range: 60 weeks to 131 weeks) was 78.8% (20.1), both *p*12 GC/kg dose (dose by ddPCR 1.0 x 10^{13} GC/kg) as the optimal biological dose for the pivotal phase 3 trial expected to start in the second half of 2021.

PrgmNr 2994 - NMD-dependent approach restores CFTR function in primary nasal cells harboring nonsense variants

View session detail

Author Block: **A. Bowling**[1], C. Merlo[2], L. Huang[3], N. West[2], S. Patel[2], G. R. Cutting[4], N. Sharma[4]; [1]Dept. of Genetic Med., Johns Hopkins Univ. Sch. of Med., Baltimore, MD, [2]Div. of Pulmonary and Critical Care Med., Dept. of Med., Johns Hopkins Hosp., Baltimore, MD, [3]Ionis Pharmaceuticals, Inc., Carlsbad, CA, [4]Johns Hopkins Univ. Sch. of Med., Baltimore, MD

Disclosure Block: **A. Bowling:** None.

Nonsense mutations remain difficult to therapeutically target, as they introduce a premature termination codon (PTC) in the mRNA resulting in degradation by nonsense mediated mRNA decay (NMD). Readthrough agents can overcome PTCs by introducing an alternate amino acid at the PTC. These therapeutics are likely to be most effective when mRNA does not undergo NMD and protein processing defects can be corrected. Additionally, antisense oligonucleotide (ASO) therapy can be used to inhibit NMD by targeting ASOs to factors in the NMD pathway, resulting in gene silencing and production of stable mRNA. Cystic fibrosis (CF), an autosomal recessive disorder caused by mutations in *CFTR,* is an ideal model system to study readthroughs, as triple combination (elexacaftor-tezacaftor-ivacaftor) can be used to correct protein defects. In this study, we tested readthrough in the presence of triple combination on human nasal epithelial (HNE) cells harboring L88X, which evades NMD. Additionally, we treated HNE cells bearing W1282X with ASOs to factors in the NMD pathway, and tested readthrough and triple combination. Cells were conditionally reprogrammed and grown at air liquid interface. CFTR function observed in WT HNE cells was 13.1Â±1.5ÂµA/cm^2 (n=12, 3-11 observations per sample). In L88X/F508del HNEs, untreated cells generated 6% of WT function. Triple combination alone restored function to 83% of WT, consistent with expected recovery of F508del CFTR. Cells treated with readthrough G418 and triple combination did not result in additional recovery of function. ELX-02, a readthrough currently in Phase II clinical trials, with triple combination exhibited a robust increase in CFTR current, up to 123% of WT. In W1282X/W1282X HNEs, untreated cells generated 1% of WT function, consistent with disease severity. Cells treated with G418 and triple combination restored function to 10% of WT. Treatment with SMG1-ASO alone did not improve function. Interestingly, when SMG1-ASO was applied in combination with G418 and triple combination, CFTR exhibited a significant increase in current, corresponding to 18% of WT. Additionally, we tested CFTR function in airway cells that are the native site of *CFTR* expression. Since splicing is essential for NMD, these cells stably express W1282X-*CFTR*-expression minigene, containing all *CFTR* exons and select abridged *CFTR* introns. Treatment with SMG1-ASO, G418, and triple combination showed significant recovery of CFTR function ($\hat{I}\Delta Isc_{treated}$=98.1ÂµA/cm^2, $\hat{I}\Delta Isc_{untreated}$=3.4ÂµA/cm^2). This work demonstrates the potential of readthrough therapeutics to target PTC-causing mutations, where NMD is evaded or can be inhibited through ASO treatment.

PrgmNr 2995 - Patient-focused drug development for a single intravenous dose of HMI-203 gene therapy in adult mucopolysaccharidosis (MPS) II, or Hunter syndrome, patients

View session detail

Author Block: **J. Haroldson**[1], C. Witalisz[1], R. A. Martin[2]; [1]Homology Med.s, Bedford, MA, [2]Private Consultant, New Florence, MO
Disclosure Block: **J. Haroldson:** Major Stockholder/Ownership Interest; Homology Medicines.
Introduction: MPS II, a rare lysosomal storage disorder caused by mutations in the IDS gene, is comprised of two subtypes: one with progressive cognitive decline and survival into the teens (neuronopathic form) and the other without cognitive impact and survival into the third or fourth decade (non-neuronopathic form). Both subtypes have debilitating peripheral manifestations. HMI-203 is an investigational AAVHSC-mediated gene therapy that provides functional copies of the IDS gene. Preclinical data in the MPS II mouse model support its potential to address both the peripheral and CNS manifestations of MPS II. Ahead of its planned first-in-human clinical trial in adults with MPS II in 2021, Homology Medicines sought to better understand the burden of disease directly from patients to inform meaningful clinical outcomes and ensure a patient-focused drug development plan. Recent focus by other drug developers is on young neuronopathic patients, with the hope of preventing or reversing the cognitive delay that enzyme replacement therapy (ERT) cannot address. The disease burdens for all MPS II patients are significant, and in the adult population, the peripheral aspects of the disease are poorly understood, and therefore have gone unaddressed despite years of weekly ERT. **Methods**: We conducted 1:1 interviews with 7 MPS II adults using a standardized questionnaire to collect qualitative data to understand the most burdensome aspects of the disease and perceptions about current and future therapies. **Results**: Median age of patients was 26 (range 24-36) years. Most Burdensome Daily Symptoms Experienced by MPS II Adults: Limited mobility and range of motion 7/7; Pain 6/7; Hearing loss 6/7; Difficulty walking or standing 3/7. Current Therapies to Manage MPS II Symptoms: ERT 7/7; Surgeries 7/7; Supportive therapies (PT/OT, chiropractic) 6/7; Hearing aids 4/7; Cardiovascular 4/7
All patients received weekly ERT and cited it as their most beneficial therapy but noted its expense and the time and inconvenience required for weekly infusions. Patients wished that ERT more adequately addressed symptoms including pain, range of motion, hearing loss and chronic fatigue. All patients desired a potential one-time gene therapy that could alleviate the burden of weekly ERT and provide at least equal therapeutic benefit. **Conclusions**: MPS II patients are burdened by weekly ERT infusions and disease manifestations that ERT does not address. Based on the results of this study of MPS II patient experiences, Homology aims to incorporate patient-focused endpoints into its Phase 1/2 clinical evaluation of HMI-203 including the potential to discontinue ERT.

PrgmNr 2996 - Somatic CAG expansion in Huntington's disease is dependent on the MLH3 endonuclease domain, which can be excluded via splice redirection

View session detail

Author Block: **R. Mouro Pinto**[1,2,3], J. Roy[1,2,4], A. Vitalo[1,2], M. Andrew[1], E. Mota-Silva[1], M. Kovalenko[1], Z. Burch[1], A. Nhu[1], P. Cohen[5], E. Grabczyk[4], V. Wheeler[1,2], R. Mouro Pinto[1,2,3]; [1]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [2]Dept. of Neurology, Harvard Med. Sch., Boston, MA, [3]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Dept. of Genetics, Louisiana State Univ. Hlth.Sci. Ctr., New Orleans, LA, [5]Ctr. for Reproductive Genomics, Dept. of BioMed. Sci., Cornell Univ., Ithaca, NY

Disclosure Block: **R. Mouro Pinto:** None.

Somatic expansion of the CAG repeat tract that causes Huntington's disease (HD) is thought to contribute to the rate of disease pathogenesis. Therefore, factors influencing repeat expansion are potential therapeutic targets. Genes in the DNA mismatch repair pathway are critical drivers of somatic expansion in HD mouse models. Here, we have tested, using genetic and pharmacological approaches, the role of the endonuclease domain of the mismatch repair protein MLH3 in somatic CAG expansion in HD mice and patient cells. A point mutation in the MLH3 endonuclease domain completely eliminated CAG expansion in the brain and peripheral tissues of a HD knock-in mouse model (*Htt*^{Q111}). To test whether the MLH3 endonuclease could be manipulated pharmacologically, we delivered splice switching oligonucleotides in mice to redirect *Mlh3* splicing to exclude the endonuclease domain. Splice redirection to an isoform lacking the endonuclease domain was associated with reduced CAG expansion. Finally, CAG expansion in HD patient-derived primary fibroblasts was also significantly reduced by redirecting *MLH3* splicing to the endogenous endonuclease domain-lacking isoform. These data indicate the potential of targeting the MLH3 endonuclease domain to slow somatic CAG repeat expansion in HD, a therapeutic strategy that may be applicable across multiple repeat expansion disorders.

PrgmNr 2997 - Sorbitol reduction in *SORD* deficient fibroblasts and iPSC-derived motor neurons using an antisense oligonucleotide (ASO) treatment strategy

View session detail

Author Block: **J. Medina**[1], A. P. Rebelo[2], C. Yanick[1], J. Yue[1], R. Shiekhattar[1], M. Saporta[1], S. Zuchner[2]; [1]Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Univ Miami, Miami, FL
Disclosure Block: **J. Medina:** None.
Recent findings have concluded biallelic mutations in *SORD*, whose gene product sorbitol dehydrogenase is the second enzyme in the polyol pathway, as the most frequent form of recessively inherited neuropathy. The polyol pathway is a two-enzyme mechanism activated in high glucose conditions, converting glucose to fructose. Aldose reductase, coded by *AKR1B1*, reduces glucose to sorbitol, which is then oxidized by sorbitol dehydrogenase into fructose. Without a functional form of sorbitol dehydrogenase, sorbitol accumulations are thought to be a catalyst for axonal damage evidenced by changes in cellular osmolarity, oxidative stress and decreased NADPH levels. Our study aims to develop genetically based treatments for *SORD* deficient cells by targeting *AKR1B1* RNA transcripts and preventing sorbitol accumulations. Using an antisense oligonucleotide (ASO) approach, we have achieved up to 80% reduction in *AKR1B1* gene product in patient fibroblasts. Currently, ASO efficacy is being investigated in an iPSC-derived motor neuron model system. Various ASO modifications and length are currently being investigated as we optimize knockdown efficiency and ASO treatment dose in both control and patient fibroblasts. *AKR1B1* transcript knockdown efficiency has been analyzed using standard western blot and RT-PCR. Intracellular sorbitol measurements are being analyzed using D-sorbitol colorimetric assay. Preliminary findings suggest a significant reduction in *AKR1B1* transcripts leading to a reduction in intracellular sorbitol accumulations. Current studies are investigating potential off-target ASO effects using comparative RNA sequencing of post-treatment patient fibroblasts and iPSC-derived motor neurons.

PrgmNr 2998 - Towards precision genome editing in an inborn error of polyamine metabolism

View session detail

Author Block: **O. Akinyele**, A. Munir, M. Johnson, J. C. Zhang, A. Arockiaraj, C. V. Land, D. Kemaladewi; Univ. of Pittsburgh, PITTSBURGH, PA
Disclosure Block: **O. Akinyele:** None.
The polyamines putrescine, spermidine and spermine, are ubiquitous polycationic molecules crucial for gene expression, signal transduction and cell proliferation. Abnormal accumulation of polyamines due to mutations in *SMS* gene encoding spermine synthase protein causes Snyder-Robinson syndrome (SRS), a rare X-linked recessive disorder which manifests as mental retardation, thin habitus, and low muscle tone (hypotonia) with no available treatments. Here, we aim to study the molecular mechanisms underlying the pathophysiology and develop a therapeutic strategy for SRS. First, we characterize a novel SRS mouse model carrying a missense mutation in the *Sms* gene, leading to a glycine-to-serine substitution (G56S) in the SMS protein. We observe a complete loss of SMS transcript and protein levels in the G56S brain, skeletal muscles, and liver. Subsequently, we quantified the level of tissue polyamines using mass-spectrometry and observed a significant increase in the spermidine content in the G56S skeletal muscles and brain. Our data so far indicates that the G56S mice recapitulate the molecular and biochemical signatures found in patients. We are currently employing a combination of imaging (MRI, micro-CT), functional (grip strength, open field, Morris water maze, fear conditioning), and histopathological techniques to evaluate the structural and functional abnormalities in the brain, bone, and skeletal muscles. In parallel, we are testing the feasibility of Prime Editing technology, the newest addition to the CRISPR/Cas toolbox, to correct the causative mutation and restore the polyamine levels. Successful completion of this project could open door to precision gene therapy for inherited diseases for which no treatment is currently available.

PrgmNr 2999 - A deeply sequenced public resource of diverse human genomes

View session detail

Author Block: **Z. Koenig**[1,2], M. T. Yohannes[1,2], J. K. Goodrich[3,2], M. Wilson[4,5], G. Tiao[4], L. Nkambule[1,2], X. Zhao[6,3,7], S. Hao[5,3], N. Sahakian[4], M. E. Talkowski[6,1,3,7], H. Brand[3,7], K. J. Karczewski[1,3,2], M. J. Daly[1,2,3,8], E. G. Atkinson[9,1,2], A. R. Martin[1,2,3]; [1]Stanley Ctr. for Psychiatric Res., The Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [3]Program in Med. and Population Genetics, The Broad Inst. of MIT and Harvard, Cambridge, MA, [4]The Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Massachusetts Gen. Hosp., Boston, MA, [6]Dept. of Neurology, Massachusetts Gen. Hosp., Boston, MA, [7]Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [8]Inst. for Molecular Med. Finland, Helsinki, Finland, [9]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX

Disclosure Block: **Z. Koenig:** None.

Diverse populations are often excluded from genomic studies due to a lack of resources supporting their analysis. The 1000 Genomes Project (1kGP) and Human Genome Diversity Project (HGDP) are among the most valuable genomic resources because of the breadth of global diversity they capture and their open sharing policies that allow release of unrestricted individual-level data. These two resources have only recently been sequenced to high coverage and have not been well harmonized in the past. 1kGP is a larger and more widely used dataset containing related individuals while HGDP spans a greater depth of diversity, though with fewer individuals from each population. HGDP therefore fills some major geographic gaps not represented in 1kGP, for example in the Middle East, Oceania, parts of sub-Saharan Africa and the Americas.

In this study, we harmonized a set of over 4,000 recently sequenced high quality whole genomes from HGDP and 1kGP (mean depth across populations=33X). We jointly called variants consisting of SNPs, indels, and SVs. After comprehensive QC leveraging the gnomAD filtering pipeline, we identified more than 155 million high-quality variants. As expected, African populations have significantly more genetic variants than out-of-Africa populations (6.1M versus 5.3M variants/individual). We ran principal components analysis (PCA) across this diverse set of populations at the global and subcontinental level. To allow projection of cohorts onto a unified reference PC space, we developed a pipeline for sharing our PC loadings to align biobanks that cannot share individual data. This resource is already being used in the COVID-19 Host Genetics Initiative and Global Biobank Meta-analysis Initiative consisting of at least 20 biobanks. We are also phasing this callset to serve as a haplotype resource for use in phasing and imputation pipelines, including in GWASpy, a cloud-based, open-source, scalable GWAS pipeline that we developed (https://github.com/atgu/GWASpy).

In sum, we demonstrate substantial added value from this joint-called dataset compared to prior resource use via variant intersection. Alongside releasing this dataset without restriction, we are providing tutorials for conducting many of the most common quality control steps and analyses with these data in a scalable compute setting. This jointly called reference panel contains individual level data from a globally-representative set of individuals and will serve as a key resource to support research of diverse ancestry populations. Including more diverse populations in statistical genetics leads to equitable genomic studies which contribute to increased global health equity.

PrgmNr 3000 - A web-based application interface for UK Biobank phenotype exploration

View session detail

Author Block: **M. Ryals**, E. Dalby, M. Permenter, L. Pi, J. Sabourin, E. Wagner, L. Hou, L. Li; BioStat Solutions, LLC, Frederick, MD
Disclosure Block: **M. Ryals:** Salary/Employment; BioStat Solutions, LLC.
The UK Biobank (UKBB) is a rich resource of phenotypic and genetic information for approximately 500K participating individuals. However, the raw tab-formatted phenotype data that is provided by UKBB requires extensive processing to be used in genetic association or other downstream analyses. Without prior programmatic experience, it can be difficult to extract meaningfully-coded phenotypes from the raw data. Examples of the mapping required include processing the phenotype field IDs, subject-level coding, repeat visits, and subject coding indices for a given phenotype field. These formatting steps require frequent referral to UKBB source material. To address this issue, we have created an R-Shiny application interface that can easily query, assess, and extract desired phenotypic information without the need for programmatic querying or running external code. The application allows a user to easily select fields at the index- or instance-level for available phenotypes and will then export an analysis-ready table for downstream genetic associations or phenotype-level studies. In addition, the application has a user-friendly method for viewing and querying UKBB disease-mapping rules for the ICD9, ICD10, primary care, and other available coding schema used on the UKBB backend for creating and refining disease phenotypes. This R-Shiny application can be easily expanded to include more data, such as disease affection status defined using the PheCode mapping recommendations , and medication usage grouped based on the Anatomical Therapeutic Chemical (ATC) Classification System. By providing an accessible interface for phenotype extraction, we hope to allow a wide audience of researchers to easily navigate and process the extensive bank of knowledge in UKBB for further analyses.

PrgmNr 3001 - An Extensible Prototype for MultiOmic Clinical Reporting

View session detail

Author Block: **B. Busby**[1], A. Al Khleifat[2], R. K. Kesharwani[3], A. Das[4], N. Giangreco[5], A. Ma[1], J. Kataria[6], A. Faranda[7], V. Chander[8], K. Pagel[9], S. Volpe[10], R. Funahashi[11], J. Kubica[12], Y. Yang[13], A. Guo[14], C. Lo[15], D. Enoma[16], A. Chander[17], M. Tandon[14], K. Narsinh[10], A. Nadkarni[14], W. Zhou[18], J. Monlong[19], A. Yadav[20], J. L. Smith[21]; [1]DNAnexus, Mountain View, CA, [2]Kingâ□□s Coll. London, London, [3]Baylor Coll. of Med., Houston, TX, [4]MIODx, San Francisco, CA, [5]DNAnexus, Columbia University, NY, [6]DNAnexus, MedGenome, India, [7]Univ. of Delaware, Newark, DE, [8]Baylor Coll. of Med. Human Genome Sequencing Ctr., Houston, TX, [9]Johns Hopkins Univ., Baltimore, MD, [10]Columbia Univ., New York, NY, [11]Univ. of Pittsburgh, Pittsburgh, PA, [12]Univ. of Warsaw, Warsaw, Poland, [13]Univ. of North Carolina, Chapel Hill, NC, [14]Carnegie Mellon Univ., Pittsburgh, PA, [15]Natl. Inst. of Genetics, Mishima City, Japan, [16]Port Harcourt, Nigeria, [17]Harvard Med. Sch., Boston, MA, [18]George Mason Univ., Fairfax, VA, [19]Univ. of California Santa Cruz, Santa Cruz, CA, [20]Academy of Scientific and Innovative Res., Ghaziabad, India, [21]Fred Hutchinson Cancer Res. Ctr., Seattle, WA

Disclosure Block: **B. Busby:** None.

Documenting the relative contributions of annotated expressed variants, structural variants, T cell receptors and variant interactions will be critical for subtyping disease states in the clinical context. As such, we have integrated multi-omics data into easily digestible comprehensive clinical reports. For each data type we have vignettes that are presented to clinicians and patients for the 5 most significant findings from that data type, as well as large scale data (in some cases genome wide) that is computationally accessible (and visually from R markdown or shiny apps) through an SQL database. For clinical researchers and healthcare professionals interested in mining and combining large scale datasets, we have also formatted these data types into the Observational Medical Outcomes Partnership (OMOP) tables. This should allow data sharing with a variety of electronic medical record systems, especially noting new initiatives by Google and others. All of this work is easily accessible on Github and some complex pipelines are often available in Workflow Description Language. We welcome individual and institutional collaborators interested in this extensible prototype for multi-omic clinical reporting.

PrgmNr 3002 - Characterizing repeat expansion variation in the Undiagnosed Disease Network cohort

View session detail

Author Block: **S. Fazal**[1], M. C. Danzi[2], U. Undiagnosed Disease Network[1], S. Zuchner[3]; [1]Univ. of Miami, Miami, FL, [2]Dakota Dunes, SD, [3]Univ Miami, Miami, FL
Disclosure Block: **S. Fazal:** None.
Expansions of short tandem repeats (TRs) are responsible for over 40 known diseases, most of which primarily affect the nervous system. We hypothesize these represent only a fraction of the pathogenic repeat expansions that exist and that they may be responsible for explaining a proportion of the missing heritability of rare monogenic diseases. ExpansionHunter Denovo (EHDn) was used to identify large TRs genome-wide in a control cohort of 2,504 genomes from the 1000 Genomes Project. This control TR profile was compared to individual disease genomes in an outlier analysis pipeline to identify TRs that are rare and expanded in cases. After processing and filtering through this pipeline, each case results in a list of candidate sites, which may be pathogenic. We processed ~2500 genomes from the Undiagnosed Disease Network (UDN) through this outlier pipeline. We found that 47% of genomes have at least 1 candidate TR with a mean of 1.5 candidate TRs per sample and a maximum of 14 candidate TRs per sample. We explored whether any known disease-causing repeat expansion loci were part of these candidate sites and found that 162 genomes were flagged for having a potentially pathogenic repeat expansion in one of nine known loci. Many of these are either heterozygous for a recessive condition or the specific sizes do not cross the known pathogenic threshold. We plan to further validate these results using repeat primed PCR (rpPCR) or gel electrophoresis. Finally, we found 339 novel candidate TRs in this cohort, 60% of which are observed in at least two genomes. These are excellent candidates for pathogenicity, and we are developing pipelines to further filter these down based on mechanistic characteristics and similarity of patient phenotypes. Importantly, our results will help to close the diagnostic gap in rare unsolved diseases.

PrgmNr 3003 - Development of a comprehensive, locus specific, database for ENPP1 Deficiency (generalized arterial calcification of infancy/autosomal recessive hypophosphatemic rickets (GACI/ARHR2) to clarify the clinical relevance of variant data

View session detail

Author Block: **C. Nester**[1], G. Khursigara[1], C. R. Ferreira[2], L. M. Chunn[3], S. A. Mercurio[3], M. J. Kiel[3]; [1]Inozyme Pharma, Boston, MA, [2]Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [3]Genomenon Inc., Ann Arbor, MI

Disclosure Block: **C. Nester:** Salary/Employment; Inozyme Pharma.

Ectonucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1) is a critical enzyme in the biochemical pathway that produces extracellular pyrophosphate (PPi) and adenosine, potent inhibitors of mineralization and neointimal proliferation, respectively. Biallelic loss of function variants in the ENPP1 gene cause ENPP1 Deficiency, a rare disorder characterized by pathological mineralization, arterial stenosis, and morbidities due to cardiovascular, pulmonary, skeletal, neurological and hearing complications. ENPP1 Deficiency is associated with 40-60% mortality early in life, and over 80% of surviving patients will develop hypophosphatemic rickets (ARHR2) or osteomalacia by age 25. Accurate and timely diagnosis is challenged by the rarity of the disease and heterogeneity of its clinical presentation. Differentiating ENPP1 Deficiency from other diseases is further complicated by the absence of a dedicated database of known *ENPP1* variants and associated clinical data.

Using a novel approach to systematic curation of genetic evidence, we have identified and analyzed all previously reported cases of GACI/ARHR2 and collated and interpreted all associated genetic variants in *ENPP1*. This technique combines automated indexing of medical literature with aggregation of population frequency databases and variant prediction algorithms followed by manual annotation and curation of this information. In total, 2,333 articles were reviewed, revealing 89 unique *ENPP1* variants identified in over 100 patients, including patients not yet published. Each variant was interpreted according to ACMG guidelines. At the time of initial analysis, 56 of these variants were classified as pathogenic/likely pathogenic (P/LP), representing a 107% increase in P/LP *ENPP1* variants documented in ClinVar. Of the P/LP variants, 59% (33/56) were missense variants, the majority of which occurred in the phosphodiesterase domain of ENPP1 (20/33). Each patient was annotated with a detailed description of their phenotypic presentation and clinical outcome. Among the patients identified with P/LP variants, the most observed phenotypes included arterial calcification, hypertension, hearing impairment, and cardiomegaly. Our analysis revealed substantial heterogeneity in disease severity, even among patients with the same variant.

This comprehensive database of *ENPP1* variants will increase the diagnostic yield of genetic testing. Establishing a locus-specific variant database and disease-specific patient database provides an important tool for the scientific and patient community to better understand this rare disease and identify novel therapeutic options.

PrgmNr 3004 - Examination of clinical features among 200,632 UK Biobank participants harboring deleterious germline variants in *NF1* and *SPRED1*

View session detail

Author Block: **A. Pemov**, J. Kim, D. Stewart; Natl. Cancer Inst., Rockville, MD
Disclosure Block: **A. Pemov:** None.
Background. Large-scale exome sequencing data linked to electronic health records (EHR) and broad ascertainment of participantsâ clinical phenotypes provides yet another opportunity to investigate genotype-phenotype correlations, improve the estimates of prevalence and penetrance of Mendelian disorders, and ultimately, advance disease prognosis, risk stratification and care for patients. **Methods.** Utilizing a genome-first approach, we identified deleterious variants in *NF1* and *SPRED1*, the causative genes in neurofibromatosis type 1 (NF1) and Legius syndrome (LS), in the exomes of 200,632 UK Biobank (UKBB) participants. We grouped deleterious variants in two tiers. Tier 1 included pathogenic/likely pathogenic variants as classified in ClinVar; and tier 2 included all remaining loss-of-function variants (nonsense, frameshifting indels and canonical splice sites). We then examined EHR for the participants carrying deleterious variants in these two genes. **Results.** We identified 43 and 25 participants carrying tier 1 and tier 2 deleterious *NF1* variants, respectively. In tier 1 and tier 2 groups, 14 and 7 participants had a âneurofibromatosisâ diagnosis in their EHR, respectively (14/43 tier 1 (33%), and 21/68 tier 1 and 2 combined (31%) variants). Highly suggestive features of NF1 (e.g., cafÃ©-au-lait macules (CALM), freckling, Lisch nodules, macrocephaly, learning delay, neurofibromas) were absent from the participantsâ records, including those diagnosed with âneurofibromatosisâ. For *SPRED1*, we observed 3 and 18 participants with tier 1 and tier 2 variants, respectively. None of the participants had records of an LS diagnosis. Highly suggestive clinical features, such as CALM, freckling, macrocephaly and learning delay were absent from their EHR as well. The prevalence ranges of deleterious variants in these two genes were close to the previously reported: 1/2,951-1/4,666 (*NF1*) and 1/9,554-1/66,877 (*SPRED1*). **Conclusions.** Among 200,632 exome-sequenced UKBB participants, only about one-third of those harboring clearly deleterious variants in *NF1* were diagnosed with âneurofibromatosisâ, and none of 21 participants with clearly deleterious variants in *SPRED1* were diagnosed with LS. These findings highlight difficulties that exist in identifying and diagnosing rare diseases in human populations using the EHR. Follow-up clinical evaluation of deleterious variant carriers would be necessary to establish reliable genotype-phenotype connections and refine the estimates of the prevalence and the penetrance of these disorders.

PrgmNr 3005 - Improving RNAseq Splice Junction Prioritization in Rare Disease Analysis

View session detail

Author Block: **B. Weisburd**[1,2], V. Ganesh[1,2,3,4], S. DiTroia[5,2,3], F. Aguet[1], G. Tiao[6,2], H. L. Rehm[2,1], A. O'Donnell-Luria[1,2,3]; [1]Broad Inst., Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Boston Childrenâ□□s Hosp., Boston, MA, [4]Brigham and Womenâ□□s Hosp., Boston, MA, [5]Broad Inst. of MIT and Harvard, Cambridge, MA, [6]Broad Inst MIT & Harvard, Cambridge, MA
Disclosure Block: **B. Weisburd:** None.
Over the past 5 years, multiple groups have developed tools and approaches for diagnosing rare disease cases using RNA-seq data - demonstrating that analyses of gene expression, splice junctions, and allele-specific expression can significantly increase diagnostic yield. Splice junction analyses in particular, are uniquely useful for interpreting deep intronic variants, and other variants of uncertain significance. However, there are remaining challenges with deploying and scaling splice junction analysis pipelines to rare disease cohorts.

First, in order to be considered a diagnosis, the splice junction outlier must be matched with an underlying causative genomic variant. For many outliers, no such variant is found. Second, even after parameter optimization, current splice junction prioritization approaches yield many 10s to 100s of significant outliers per sample. These two issues together lead to prohibitive amounts of manual review being necessary to consider all splice junction outliers in each case.

Here we present a pipeline that lessens the burden of downstream manual review and improves splice junction prioritization. The pipeline first aligns all RNA-seq samples to GRCh38 by running the STAR v2 aligner [Dobin 2013]. We use identical parameters to those used in the GTEx v8 release [GTEx Consortium 2020] to enable comparison of our case samples with tissue-matched GTEx controls. To do this comparison, we apply the outlier detection approaches described in [Cummings 2017] and [Mertes 2021], yielding a list of candidate splice junctions. Next, we apply Portcullis [Mapleson 2018] to filter junctions and exclude technical artifacts which decreases the candidate splice variant list by ~30%. Then, for cases where short or structural variant calls are also available from exome or genome data, we annotate the splice junctions with information about any rare variants present within 10kb of each splice junction. Finally, we re-prioritize the candidate junctions based on these additional annotations, yielding a smaller set of junctions for manual review.

We present the pipelineâ□□s accuracy based on truth data derived from previous publications and show the results from applying this pipeline to over 100 unsolved cases in a heterogeneous rare disease cohort. With this strategy, we reduce the number of splice junctions and variants requiring analysis to achieve diagnosis.

PrgmNr 3006 - Improving the imputation quality of French-Canadian genomic data

View session detail

Author Block: **J. Pelletier**[1], J-C. Grenier[2], p. raphael[3], J. Hussin[1]; [1]MontrÃ©al Heart Inst., MontrÃ©al, QC, Canada, [2]5000 rue BÃ©langer, MontrÃ©al, Canada, H1T 1C7, Montreal, QC, Canada, [3]UdeM, MontrÃ©al, QC, Canada

Disclosure Block: **J. Pelletier:** None.

Imputation in genetics is the statistical inference of unobserved genotypes. This process is influenced by variants localization, recombination, and the demographic history of the population, however, this last factor isnâ□□t generally considered in current imputation strategies. In the French-Canadian population, the complex demographic history could impact genotype imputation procedure with current reference cohorts. In this project, we test whether the use of more varied reference cohorts increases the quality of imputation in this founder population. To characterize the quality of imputation for the French-Canadian population, we used two reference cohorts: TOPMed and Haplotype Reference Consortium (HRC). The imputation is made on the 30,000 genotyped individuals from the CARTaGENE project. Our preliminary results show that the number of sites imputed with high quality is significantly higher with TOPMed compared to HRC. However, by using the same genotyping chip for Canadian sub populations from other provinces, we note that the imputation is less efficient for the French-Canadian founder population. Finally, by analyzing the imputed variants with respect to their imputation scores, we find that variants specific to the founder population were possibly wrongly attributed with a bad quality score. We therefore confirm that the use of TOPMed makes it possible to improve the imputation of genetic data from French Canadians, but there remains a gap with non-founder populations, potentially due to allele frequencies mismatch from the mean frequencies in the reference cohorts used for the imputation. Other quality criteria may be needed for imputation in the French-Canadian population and in other human founder populations.

PrgmNr 3007 - Making Discoveries with Kids First Variant Database

View session detail

Author Block: **Y. Guo**[1], J. Costanza[2], C. Botek[2], M. Brown[1], D. Higgins[1], Y. Zhu[3], B. Farrow[3], A. Heath[3], A. Resnick[3], V. Ferretti[2], Kids First Data Resource Center; [1]The Children's Hosp. of Philadelphia, Phiadelphia, PA, [2]Ctr. Hosp.ier universitaire Sainte-Justine, Montreal, QC, Canada, [3]The Children's Hosp. of Philadelphia, Philadelphia, PA

Disclosure Block: **Y. Guo:** None.

The Gabriella Miller Kids First Pediatric Research Program (Kids First) aims at facilitating researchers to uncover new insights into the biology of childhood cancer (CC) and structural birth defects (SBD), including the discovery of shared genetic pathways between these disorders. Kids First has two initiatives, which are whole genome sequencing (WGS) of biospecimens from families with CC/SBD, and establishing Kids First Data Resource. Kids First Data Resource Center (KFDRC) developed Kids First Data Resource Portal (KFDRP; https://portal.kidsfirstdrc.org/), which is a centralized platform to search, view, analyze, and identify currently accessible data from both Kids First and collaborative cohorts, incorporating omics and phenotypic information of 23 studies and 20,097 participants. Recently KFDRC released two new KFDRP components named Variant DataBase (VDB) and Variant WorkBench (VWB), enabling users to query, mangle, analyze and visualize germline genomic variants. The first release of VDB on 1/21/2021 includes ~216 million unique variants in a matrix of more than 26.5 billion individual-chromosomal position occurrences from over 5,000 participants in 6 studies. While VDB provides a quick variant summary, VWB supports scripting languages such as Python, Spark, SQL and MarkDown as in-depth analysis tools enabled by Apache Zeppelin notebooks. In addition to variant calls and phenotypic information such as Human Phenotype Ontology (HPO) terms and Mondo IDs, VWB hosts rich external variant annotations in the public domain, such as Cancer Hotspots, ClinVar, COSMIC, dbNSFP, gnomAD, TOPMed, as well as gene-phenotype links provided by OMIM, HPO, Orphanet, and the Deciphering Developmental Disorders Project. Users can also visualize analysis results in multiple chart styles, display local figures, import custom datasets as temporary query tables, and export analysis results to local drives. In an effort to screen fibroblast growth factor receptors (FGFR) genes for deleterious variants in a Kids First cohort of kidney and urinary tract defects, we identified a likely pathogenic de novo variant in FGFR3 by running a series of PySpark/SQL scripts in VWB. We first restricted gene symbol to begin with â□□FGFRâ□□ and required at least 19 out of the 20 variant consequence predictions to be damaging/deleterious/detrimental, then checked the variantâ□□s genotype status in other family members, and confirmed it is absent from any public database such as gnomAD/TOPMed/ClinVar. FGFR3 (OMIM *134934) has been linked to multiple autosomal dominant genetic disorders and two of them (#149730 and #612247) involve clinical phenotypes in the genitourinary system.

PrgmNr 3008 - NCBI ALFA Release 2for900 Million Variantsand Allele Frequencyfrom 200KdbGaPSubjects

View session detail

Author Block: **L. Phan**[1], Y. Jin[2], H. Zhang[1], Q. Wang[1], G. Shekhtman[1], D. Shao[1], R. R. Villamarin[1], M. Kimura[1], J. Wang[1], L. Hao[1], N. Sharopova[1], M. Bihan[1], A. Sturcke[1], M. Lee[1], N. Popova[1], W. Wu[1], C. Bastiani[1], M. Ward[3], B. Holmes[1], V. Lyoshin[1], K. Kaur[1], E. Moyer[1], M. Feolo[1], B. L. Kattman[4]; [1]NIH, NLM/NCBI, Bethesda, MD, [2]Natl. Library of Med., NIH, Bethesda, MD, [3]NIH Natl. Ctr. for, Bethesda, MD, [4]NCBI/NLM/NIH, Fort Collins, CO

Disclosure Block: **L. Phan:** None.

NCBI Allele Frequency Aggregator (ALFA) aims to provide the largest and most comprehensive aggregated variant datasets with allele frequency from dbGaP studies as open-access. dbGaP has over two million subjects and up to billions of variants along with thousands of phenotypes and molecular assay datasets. This unprecedented volume and the variety of data hold huge opportunities to exploring and studying genetic variations within human populations and identify genetic factors that influence health and diseases to improve diagnosis, treatment, and prevention.

ALFA Release 2 has over 900 million variants, including 300 million novel variants not in dbSNP Build 154. The data was generated from 79 dbGaP studies that included 192 thousand subjects and 5.8 trillion combined genotypes. Allele frequencies are available for 12 populations, including European, Hispanic, African, Asian, and other diverse population ancestries.

Allele frequencies are available for 86% of variants in dbSNP Build 155 (920M rs), 86% (334K rs) ClinVar small variants, and 99% (22K rs) of variants in the dbGaP GWAS catalog.

This massive amount of data is available as open-access from NCBI for variant interpretation and analysis. It is accessible by web search, FTP download, retrieval using API, and TrackHubs for genomic browsers. Please visit the ALFA homepage for more information about the project, releases, tutorials, and past presentations. ALFA website: https://www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/

PrgmNr 3009 - NIA Genetics of Alzheimerâ□□s Disease Data Storage Site (NIAGADS): 2021 Update

View session detail

Author Block: **H. Issen**[1], A. Kuzma[1], O. Valladares[1], E. Greenfest-Allen[1], C. Klamann[1], P. Gangadharan[1], Z. Katanic[1], A. Wilk[1], Y. Zhao[2], L. Qu[2], M. Moon[1], A. Lerro[1], J. Manuel[1], P. Keskinen[1], C. Thomas[1], S-Y. Chou[3], W-P. Lee[1], Y. Leung[1], A. C. Naj[2], C. J. Stoeckert Jr.[1], G. D. Schellenberg[4], L-S. Wang[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [3]Dept. of Economics, Lehigh Univ., Bethlehem, PA, [4]Univ Pennsylvania Sch Med, Philadelphia, PA

Disclosure Block: **H. Issen:** None.

Background: NIAGADS is a national genomics data repository that facilitates access of genotypic and sequencing data to qualified investigators for the study of the genetics of Alzheimerâ□□s disease (AD) and related neurological diseases. Collaborations with large consortia and centers such as the Alzheimerâ□□s Disease Genetics Consortium (ADGC), Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the Alzheimerâ□□s Disease Sequencing Project (ADSP), and the Genome Center for Alzheimerâ□□s Disease (GCAD) allow NIAGADS to lead the effort in managing large AD datasets that can be easily accessed and fully utilized by the research community.
Method: NIAGADS is supported by National Institute on Aging (NIA) under a cooperative agreement. All data derived from NIA funded AD genetics studies are expected to be deposited in NIAGADS or another NIA approved site. NIAGADS manages a Data Sharing Service (DSS) that facilitates the deposition and sharing of genomic data and association results with approved users in the neurodegenerative research community. In addition, researchers are able to freely use the NIAGADS Alzheimerâ□□s Genomics Database (www.niagads.org/genomics/) to search annotation resources that link published AD studies to AD-relevant sequence features and genome-wide annotations.
Result: As of June 2021, NIAGADS houses 76 datasets comprised of >90,000 samples including GWAS, sequencing, gene expression, annotations, deep phenotypes, and summary statistics. Qualified investigators can retrieve ADSP sequencing data with ease and flexibility through the NIAGADS DSS. As of June 2021, the ADSP and other contributing studies have completed whole exome sequencing (WES) of 20,504 samples and whole-genome sequencing (WGS) of 16,906 samples. Raw WES and WGS files, quality controlled VCF files, and phenotype data files are available via qualified access. The next round of sequencing currently underway will generate around 18,000 additional genomes to be released at the middle of 2022.
Conclusion: NIAGADS is a rich resource for AD researchers, with the goal of facilitating advances in Alzheimerâ□□s genetics research. By housing datasets from many projects and institutions, NIAGADS enables AD researchers to meet their research goals more efficiently. Datasets, guidelines, and new features are available on our website at https://www.niagads.org.

PrgmNr 3010 - Phenome-wide association analysis of rare and common variation in 455,000 UK Biobank exome sequences accessible through public data portal

View session detail

Author Block: **Q. Wang**[1], K. Carss[2], R. S. Dhindsa[1], A. Harper[2], A. Nag[2], I. Tachmazidou[2], D. Vitsios[2], S. Deevi[2], J. Okae[2], S. Wasilewski[2], S. Katherine[2], S. Petrovski[2], AstraZeneca Genomics Initiative; [1]Ctr. for Genomics Res., AstraZeneca, Waltham, MA, [2]Ctr. for Genomics Res., AstraZeneca, Cambridge, United Kingdom
Disclosure Block: **Q. Wang:** Salary/Employment; AstraZeneca.
We adopted the exome sequences from ~455K UK Biobank participants to study the contribution of both rare and common protein-coding variation to thousands of phenotypic endpoints.
We performed a gene-based collapsing PheWAS and separately a variant-level PheWAS across ~17K binary/quantitative phenotypes, evaluating a range of genetic architectures. These analyses were conducted on our cloud-based platform. Overall, we identified 1,517 gene-phenotype relationships for binary phenotypes and 1,301 gene-phenotype relationships for quantitative phenotypes that were significantly associated (p-9), reflecting 125 and 378 distinct genes, respectively. On comparing with the OMIM database, among the significant findings for binary (clinical) phenotypes, 81% are previously known, 9% are known disease genes but represent novel phenotypic associations, and 10% are novel disease genes. The variant-level analysis yielded 5,178 significant (p-9) non-synonymous variant-phenotype relationships for binary phenotypes and 34,561 significant (p-9) non-synonymous variant-phenotype relationships for quantitative phenotypes, involving 586 and 3,860 distinct genes, respectively. Interestingly, 79% (1,197/1,517) and 46% (596/1,301) of the significant gene-phenotype relationships identified in the collapsing analysis for binary and quantitative phenotypes, respectively, were not detected in the corresponding single-variant analysis, demonstrating the complementarity of the two analytical approaches.
We will illustrate AstraZenecaâ□□s use of these data to rapidly validate and inform safety profiles of potential drug targets, identify novel targets, predict drug repositioning opportunities, and study prevalence of molecular endotypes. We also introduce a data portal to visually navigate this rich resource of PheWAS summary statistics for 19K genes and across 3.2M protein-coding variants. The portal supports the search of associations by gene, phenotype and variant, data visualisations to help interpret and explore the results, and ability to download focused results for further analysis.

PrgmNr 3011 - Real-world Evidence for Human Phenotype Ontology Concept Prevalence and its Association with Genetic Disorders from Electronic Health Records

View session detail

Author Block: **C. Liu**[1], C. Ta[1], J. Havrilla[2], K. Wang[3], C. Weng[1]; [1]Columbia Univ., New York, NY, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]CHOP, Philadelphia, PA
Disclosure Block: **C. Liu:** None.
Background: The Human Phenotype Ontology (HPO) is a powerful resource for annotating and analyzing human genetic disorders. However, it is mainly curated based on expertsâ knowledge and public case reports. There is a great demand for data-driven knowledge discovery of its conceptsâ prevalence and associations with genetic disorders, supported by real-world evidence. **Methods:** We first pre-processed unstructured clinical narratives from Columbia University Irving Medical Centerâs (CUIMC) electronic health record (EHR) data prior to 2020 February and indexed them together with relevant metadata using Solr. For each HPO concept, we queried for relevant notes by including both the conceptâs syntactic variations and synonyms, except when these terms were negated or family member-related mentions by using negation and family triggers. Concepts were aggregated at the patient level to calculate each conceptâs prevalence. The âsuspiciousâ high prevalence concepts were identified, and their queries were then examined and corrected if necessary. The whole process was iterated a few times until no âsuspiciousâ concepts were found. We further generated a patient-level summary of genetic disorders by applying a similar procedure to the concepts defined in OMIM and ORPHANET. The ratios of observed-expected co-occurrence frequency between HPO concepts and genetic disorders were calculated to reflect their associations. **Results:** A manual review of the HPO prevalence found the EHR-derived HPO prevalence was consistent with HPOâs original concept prevalence annotation. Additional evaluation based on the automated mapping between a subset of HPO and ICD codes indicates our pipeline based on information retrieval using unstructured narratives reached a similar accuracy as using structured billing codes but covered more phenotypes and genetic disorders. The bulk dataset, derived from all records, contains >10,000 unique HPO concepts (>20,000 terms) based on >2 million patients. To facilitate the analysis of phenotype-disease association for various purposes, we also stratified the data into the pediatrics dataset, genetic counseling dataset, neurology dataset, and whole-exome sequencing dataset. **Conclusion:** We have derived real-world evidence-based HPO prevalence and its association with genetic disorders using CUIMC EHR. The aggregated summary statistics have been integrated into our previously developed Columbia Open Health Data (COHD) for public access through a web application programming interface (API).

PrgmNr 3012 - Tools and Resources to Improve Understanding of Rare Genetic Conditions in Newborn Screening

View session detail

Author Block: **J. Taylor**[1], A. M. Brower[2]; [1]American Coll. of Med. Genetics and Genomics, Bethesda, MD, [2]American Coll. of Med. Genetics and Genomics, Dakota Dunes, SD
Disclosure Block: **J. Taylor:** None.
Newborn screening (NBS) is an essential public health program that began with phenylketonuria (PKU) and has expanded to over 35 core and 26 secondary conditions recommended for screening by the Secretary U.S. Department of Health and Human Services. New conditions are nominated by a community of researchers, clinicians, advocates, and families from the rare disease community for consideration at the federal level by the Advisory Committee for Heritable Disorders in Newborns and Children (ACHDNC) to the Recommended Uniform Screening Panel (RUSP) at an accelerated rate. Advocates are also pursuing the addition of conditions to NBS panels at the state level as well. To help rare disease stakeholders, the Newborn Screening Translational Research Network (NBSTRN) had developed several tools and resources to navigate NBS research and clinical information.
The Newborn Screening Conditions Resource (NBS-CR) is an interactive tool that is available online at nbstrn.org/tools/nbs-cr. This tool currently features 118 conditions including the 61 conditions currently on the RUSP, 20 conditions not on the RUSP but are being screened by at least one state, and 37 conditions that NBSTRN identified as candidates. The NR-CR is a centralized resource that links to facts and statistics from many resources including: OMIM, GeneReview, MedGen, ClinicalTrials.gov, NIH RePORTER, the ACMG ACT sheets, Genetic Alliance, as well as others for both screened and candidate conditions.
NBSTRN also hosted monthly webinars to facilitate information sharing between state newborn screening programs, researchers, clinicians, advocates, and federal partners about conditions recently added to the RUSP or conditions currently apart of pilot studies. Conditions on these calls include pompe disease, mucopolysaccharidosis type I, x-linked adrenoleukodystrophy, spinal muscular atrophy, Duchene muscular dystrophy, and many more. NBS stakeholder learn about: screening assay sensitivity and specificity, incidence of the condition and how it differs between states, other disorders that may be identified by the screening method, as well as treatment and clinical management of diagnosed patients. Genetic testing has become more prominent in NBS to reduce false positive rate or to provide more information to clinicians. Recently, these webinars have discussed how NBS programs classified genetic variants and how this information effects the management of these infants.
In conclusion, NBSTRN has developed tools and resources to advance research in rare genetic diseases that affect infants and children as well as the understanding of genetic medicine.

PrgmNr 3013 - Toward a gold-standard set for SNP to gene mapping

View session detail

Author Block: **E. Fauman**[1], on behalf of ICDA Working Group 17, C. Anderson[2], L. Franke[3], X. Hu[1], G. Lettre[4,5], A. Mahajan[6], M. McCarthy[6], D. Ochoa[7,8], B. Richards[9], G. Trynka[2,7]; [1]Pfizer, Cambridge, MA, [2]Wellcome Sanger Inst., Hinxton, Cambridgeshire, United Kingdom, [3]Univ. Med. Ctr. Groningen, Groningen, Groningen, Netherlands, [4]Montreal Heart Inst, Montreal, QC, Canada, [5]Dept. of Med., Faculty of Med., UniversitÃ© de MontrÃ©al, MontrÃ©al, QC, Canada, [6]Wellcome Ctr. for Human Genetics, Nuffield Dept. of Med., Univ. of Oxford, Oxford, United Kingdom, [7]Open Targets, Wellcome Genome Campus, Cambridge, United Kingdom, [8]European Molecular Biology Lab., European Bioinformatics Inst. (EMBL-EBI), Wellcome Genome Campus, Hinxton, United Kingdom, [9]McGill Univ., Montreal, QC, Canada

Disclosure Block: **E. Fauman:** Salary/Employment; Pfizer, Inc.

Genome-wide association studies (GWAS) have been hugely successful in identifying reproducible genetic associations for a wide range of traits, biomarkers and diseases. However, our ability to derive biological and medical insights from GWAS is restricted by our limited understanding of how lead variants link to causal genes and the downstream biology. Many methods have been developed in recent years to address this so called â□□SNP-to-gene problemâ□□, but such methods require an authoritative â□□gold standardâ□□ positive control set as a starting point. No large well-curated positive control set currently exists. In response to this gap the International Common Disease Alliance (ICDA) commissioned a working group to define and populate a positive control set of high confidence causal genes (ICDA working group 17). We report here for the first time the results of this effort. We started by analyzing previous high confidence causal gene annotations, including those for cis-protein QTLs, metabolite QTLs and well-studied diabetes loci. In bridging the causal chain from SNP to trait we considered both â□□bottom-upâ□□ genomic features, which link a variant to the causal transcript, and â□□top downâ□□ biological features, which link a causal transcript to the GWAS trait. We manually reviewed thousands of loci for hundreds of traits, prioritizing features such as cell-type specific open chromatin, direct experimental evidence, animal model systems, rare disease phenotypes, rare coding variants and known drug targets. We have manually curated a set of over 2500 gene-trait positive control assertions covering over 600 metabolite traits and 250 disease and other traits. These 2500 assertions implicate a total of over 750 unique genes. Each curated assertion is annotated with a specific rationale and a link to the most direct supporting evidence. Automated annotations for each assertion provide relevant mouse knock-out phenotypes, relevant rare disease phenotypes, known missense/coding evidence and relevant drug target information. As applied to the GWAS catalog, these 2500 gene-trait assertions nominate a probable causal gene for 13% of the 190,000 genome-wide significant SNP-trait entries. This initial set of positive control genes is intended to provide a framework for accumulating and annotating additional high confidence causal gene assignments. Studying the top-down and bottom-up features of these positive control genes should help advance the field in the development of improved SNP-to-gene methods, building out the causal chain from GWAS variants to observed phenotypes, thereby advancing our knowledge of biology and pathobiology.

PrgmNr 3014 - An individual with craniosynostosis, hypertelorism, and imperforate anus with normal cognition caused by a 7.4 Mb duplication on chromosome 6

View session detail

Author Block: **V. Keesee**[1], E. J. Bhoj[2], D. Li[3], H. Hakonarson[4], A. Sobering[5]; [1]SGU, St. George's, Grenada, [2]Philadelphia, PA, [3]CHOP, Philadelphia, PA, [4]Children S Hosp. of Philadelphia, Philadelphia, PA, [5]St. George's Univ., Great River, NY
Disclosure Block: **V. Keesee:** None.
We present a 29-year-old man who has normal cognition but had imperforate anus and craniosynostosis as an infant. A head structural malformation was first observed prenatally when ultrasound showed premature closure of the coronal and metopic sutures. He has hypertelorism and deformation of the orbital wall of the right eye; when right eye is neutral, it is a bit lower than it should be. His right audocanal is shorter than the left. He is prone to a right shoulder dislocation. All developmental milestones were on time, and he has excelled academically. He sought genetic diagnosis because of his appearance. Through proband-only exome sequencing and copy number variant (CNV) analysis, we identified an approximate 7 Mb tandem duplication of 6p.22.1 - 22.3. Cytogenomic microarray analysis (CMA) using a GSA-SNP array showed a 7.4 Mb duplication on chromosome 6. The CMA probes allowed estimation of the approximate breakpoints to be within chromosomal coordinates 20,582,144 to 28,040,581.This CNV was not detected by karyotype which was done when he was an infant. The duplicated segment includes many genes including a histone supercluster, and *HFE*. This individual has phenotypic overlaps with *TWIST*, *MID1*, and *SPECC1L*. Prior to exome sequencing we considered Opitz/GBBB as part of our differential diagnosis. To our knowledge, the individual we describe has a novel chromosomal abnormality, which has not been previously described.

PrgmNr 3015 - Development of a Genetic Diagnostic Algorithm for Individuals with Split Hand Foot Malformation

View session detail

Author Block: **Y. Wang**[1], H. Alrukban[1], L. Domeshek[2], E. Ho[2,3], L. Dupuis[1], G. Borschel[2], R. Mendoza-Londono[1]; [1]Div. of Clinical and Metabolic Genetics, The Hosp. for Sick Children and Univ. of Toronto, Toronto, ON, Canada, [2]Div. of Plastic and Reconstructive Surgery, The Hosp. for Sick Children, Toronto, ON, Canada, [3]Rehabilitation Sci. Inst., Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **Y. Wang:** None.
Background: Split hand foot malformation (SHFM) is a distal limb defect that affects the central rays of the limb and accounts for up to 17% of limb reduction defects. SHFM can occur in isolation or as part of a genetic syndrome with considerable genetic heterogeneity. Mechanisms of the disease include monogenetic and chromosomal, which leads to significant challenges for genetic diagnosis and counselling. Also, no current genetic diagnostic algorithm is available.
Methods: Through a retrospective chart review, we analyzed a cohort of 34 patients with SHFM who underwent genetic tests and treatment at the Hospital for Sick Children between 1970 and 2016. To the best of our knowledge, this is currently the largest cohort of SHFM patients reported in the literature.
Results: In our cohort, about half of the patients had involvement of both upper and lower limbs (47%, 16/34), followed by upper limbs only (38%, 13/34) and lower limbs only (15%, 5/34), most with right-sided involvement (72.7%). Three patients (9%) had SHFM with long bone involvement, of which two involve the upper limbs and one affects the fibulae. Complex SHFM with associated anomalies, including skeletal, ectodermal, neurological, gastrointestinal, and urogenital systems, was identified in 38% (13/34) of the cases. About a quarter of patients (24%, 8/34) had an affected first or second-degree relative. Of the 29 patients with upper extremities involvement, over half (55%, 16/29) had good functions and did not require surgical treatment. Of the 18 patients who underwent at least one genetic test, including karyotyping, chromosomal microarray and single-gene testing (TP63, DLX5, WNT10B), 5 cases (28%, 5/18) reached a genetic diagnosis (one patient with 10q24.32 duplication, one patient with 7q21.3q32 inversion, and three patients with a pathogenic TP63 variant).
Discussion: Combining the data from our cohort and the available SHFM cases reported in the literature, we propose an algorithm for the genetic diagnosis of SHFM. We categorize the SHFM into four categories: isolated SHFM, SHFM with preaxial involvement, SHFM with long bone involvement, and complex SHFM. For each category, we set the priority for karyotyping, chromosomal microarray, and single-gene testing base on the clinical presentations and family history. Whole-genome sequencing might be a future diagnostic and gene discovery tool to reveal both single-gene and structural variants, especially for complex SHFM. We believe the proposed algorithm will lead to a better understanding, management, and counselling of individuals with SHFM and their families by clinical geneticists and plastic surgeons.

PrgmNr 3016 - Missense variants affecting the actin-binding domains of *PLS3* cause X-linked congenital diaphragmatic hernia and body wall defects

View session detail

Author Block: **F. High**[1], F. Petit[2], M. Longoni[3], J. Wells[4], R. Maser[4], M. Dysart[1], H. T. M. Contreras[1], F. FrÃ©nois[5], E. Bogenschutz[4], B. R. Pober[6], R. D. Clark[7], P. F. Giampietro[8], M. Loscertales[1], X. Ai[1], H. Brand[1], D. J. McCulley[9], L. Qiao[10], R. Hernan[11], J. Wynn[12], T. M. Scott[13], D. Calame[14], Z. Coban-Akdemir[15], H. Yonath[16], J. R. Lupski[17], Y. Shen[18], W. K. Chung[19], D. A. Scott[20], C. J. Bult[4], P. Donahoe[1];
[1]Massachusetts Gen. Hosp., Boston, MA, [2]CHRU Lille, Lille, Nord, France, [3]Tempus, Chicago, IL, [4]The Jackson Lab., Bar Harbor, ME, [5]CHU Lille, Lille, France, [6]Mass Gen Children's Hosp, Boston, MA, [7]Loma Linda Univ Med Ctr, Loma Linda, CA, [8]Robert Wood Johnson Sch. of Med., New Brunswick, NJ, [9]Univ. of San Diego, San Diego, CA, [10]Columbia Univ. Irving Med. Ctr., Fort Lee, NJ, [11]CUIMC, New York, NY, [12]New York Presbyterian Columbia, New York, NY, [13]Brigham Young Univ., Provo, UT, [14]Houston, TX, [15]Baylor Coll. of Med., Houston, TX, [16]Sheba Med Ctr, Ramat Gan, Israel, [17]Baylor Col Med, Houston, TX, [18]Columbia Univ., New York, NY, [19]Columbia Univ, New York, NY, [20]Baylor Coll. Med, Houston, TX
Disclosure Block: **F. High:** None.

Congenital diaphragmatic hernia (CDH) is a relatively common and genetically heterogeneous structural birth defect associated with high mortality and morbidity. We describe eight unrelated families with a novel X-linked condition characterized by diaphragm defects, variable anterior body wall anomalies, and/or facial dysmorphism. Using linkage analysis and whole exome or whole genome sequencing, we identified novel missense variants in the actin binding domains of plastin 3 (*PLS3*), a gene encoding an actin bundling protein, that co-segregate with disease in all families. Loss-of-function variants in *PLS3* have been described previously in association with X-linked osteoporosis. To address these seemingly disparate clinical phenotypes, we performed *in silico* protein modeling and cellular overexpression experiments, which suggest that the affected residues in individuals with CDH are important for actin binding and result in disorganization of the actin cytoskeleton and a reduction in normal actin stress fiber formation. A mouse knock-in model of a variant identified in one of the families, p.W499C, shows partial perinatal lethality and recapitulates the key findings of the human phenotype, including diaphragm and abdominal wall defects. Both the mouse model and one surviving adult patient with a *PLS3* variant were observed to have increased, rather than decreased, bone mineral density. Together, these clinical and functional data in human and mouse reveal that specific missense variants affecting the actin binding domains of *PLS3* may have a gain-of-function effect and cause a new Mendelian disorder.

PrgmNr 3017 - Multiple independent genetic syndromes in a single patient

View session detail

Author Block: **D. Watson**; CHOP, PHILADELPHIA, PA
Disclosure Block: **D. Watson:** None.
We report on an 18-year old female patient with over 662 Mb regions of homozygosity presenting clinically with intellectual disability, ataxia, schizophrenia symptoms, retinal dystrophy, moderate to severe progressive sensorineural hearing loss (SNHL), congenital hypothyroidism, cleft mitral valve with mild mitral valve regurgitation, and some dysmorphic features. Exome analysis uncovered the following variants: 1) two variants in BBS6 potentially causative for Bardet-Biedl Syndrome 6; 2) a homozygous, known pathogenic variant in the stereocilin (STRC) gene associated with non-syndromic deafness; 3) a homozygous variant in dual oxidase 2 (DUOX2) gene associated with congenital hypothyroidism; and 4) a single variant in the troponin T2 (TNNT2) cardiac type gene related to a cardiomyopathy and manifestation of cleft mitral valve with mild mitral valve regurgitation. This patient was found to be the product of a consanguineous union of an unspecified but apparent first-degree relationship, explaining the multiple independent inherited findings. In summary we report an 18-year Indian girl with a variety of clinical symptoms that are not consistent for only one genetic condition, which is reinforced by the presence of multiple regions of homozygosity consistent with an apparent first- degree relationship. Genetic variants were found in four genes BBS6/MKKS, STRC, DUOX2, and TNNT2, not all of them were confirmed pathogenic but highly suspicious to explain the clinical findings. This case report is a good example of the use of research exome in the scenario of reported negative clinical exome.

PrgmNr 3018 - Seckel-6 Syndrome: A new severe phenotype

View session detail

Author Block: **H. Goodwin**, N. Hauser; INOVA Children's Hosp., Fairfax, VA
Disclosure Block: **H. Goodwin:** None.
Seckel syndrome comprises 6 different genes, each with a slightly different phenotype but all include the characteristic pattern of microcephaly and dwarfism. The only known cases of Seckel-6 syndrome were four cousins born to parents in a consanguineous Pakistani family. These cousins were homozygous for c.129G>A; p.W43X.They displayed significant microcephaly at birth with FOC -4SD to -6SD and short stature, heights were noted at between -2SD to -4SD. Developmentally, the cousins had speech delay but were able to talk by age 3 years of age, they had mild to moderate cognitive impairment, but their motor skills were normal. Our patient displayed significantly different and more severe features than what has been previously described. Our patientâ□□s case may contribute to an expansion of the phenotype for this condition.
The genetic testing done on our patient included a karyotype of 46,XX, FISH negative for trisomy 18, and a normal microarray, other than 8% homozygosity. The Comprehensive Primordial Dwarfism Sequencing Panel (U of Chicago) revealed a homozygous deletion in the CEP63 gene (c.712._730del; p.Val238Serfs*12).Our patientsâ□□ measurements at birth showed a birth weight of 1275g (-5.5 SD), length 38.5cm (-5.7 SD), and HC of 24.5cm (-7.9 SD). Her exam was notable for down slanting palpebral fissures, posteriorly rotated ears with small dysplastic ear lobes and a prominent anti-helix, a large prominent nose, microretrognathia, a short neck with redundant skin, shield shaped chest, ambiguous genitalia with a prominent vaginal/hymen tag, brachydactyly with contractures at the elbows and knees bilaterally, overlapping digits, widely spaced toes, and her scalp exam showed disorganized hair whorls.
She reached the 5-10th percentile for weight where she remained the majority of her life. She remained neurologically impaired, unlike the previously described children. She developed seizures by age 2 years that required anti-epileptics. Later, she became more aggressive with self-injurious behavior requiring medication. Neuroimaging was significant for a severe brain malformation with essentially absent normal cerebral formation. She developed hydrocephalus requiring a VP shunt and needed a cranioplasty for craniosynostosis by age 3 years. She passed away from complications of an infection at age 6 years.
Here we describe a case of Seckel-6 syndrome, with the typical findings of severe microcephaly and short stature, but several additional features not yet noted in the literature. It is possible our patientsâ□□ additional features could be due to another underlying genetic condition, or she may have demonstrated an expansion of the phenotype for Seckel-6.

PrgmNr 3019 - The Expanding Phenotype of WAGR Syndrome: Reconceptualization from a Syndrome to a Spectrum Disorder

View session detail

Author Block: **J. Kalish**, International WAGR Syndrome Association (IWSA), K. A. Duffy; Children's Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **J. Kalish:** None.
WAGR syndrome is a disorder caused by a deletion affecting chromosome 11p13 and classically characterized by the presentation of the classic features Wilms tumor (WT), Aniridia, Genitourinary anomalies (GU), and Range of developmental delays. In addition to the classic WAGR phenotypic features, there has also been an associated risk for obesity and kidney failure. Various subgroups of patients with WAGR syndrome have been described, with the most common: â□□AGRâ□□ triad, to describe patients without WT development; and â□□WAGROâ□□ or â□□WAGR plusâ□□ to describe patients affected by obesity.The majority of previous investigations into WAGR syndrome have focused on specific features of the disorder, and only one previous study has performed a widespread evaluation into the clinical issues faced by patients with WAGR syndrome. In this study, we utilized the WAGR Syndrome Patient Registry to perform a comprehensive evaluation of self-reported health issues in 91 participants affected by WAGR syndrome. The main objectives were to evaluate the prevalence of classic phenotypic features, as well as identify any common health issues or features in the cohort.The classic WAGR features were common in the cohort, and previous associations with obesity and kidney failure were confirmed in this cohort. A high rate of issues beyond the classic WAGR phenotype were observed, which suggests the disorder would be better conceptualized as a spectrum disorder. In addition to the classic WAGR features, high rates of obesity, hypertension, and chronic kidney disease (CKD) were found, and cardiac and pulmonary features were common. Given these associations, patients with WAGR spectrum can be considered at risk for adverse cardiometabolic health and CKD in addition to the established WT risk. Internal GU anomalies and congenital anomalies affecting the kidney and/or urinary tract (CAKUT) were common in addition to external GU anomalies, suggesting a potential association between CAKUT and WAGR spectrum.In summary, we propose the concept of â□□WAGR Spectrumâ□□ which can be considered an umbrella term to describe the constellation of phenotypes presenting in patients affected by chromosome 11p13 deletions. Our results suggest a wide array of phenotypes and clinical issues in the largest cohort evaluated in this patient population and further suggest the complete phenotypic spectrum of WAGR syndrome has likely been underrepresented in past cohorts. These observations will augment the diagnosis and management of future patients with WAGR Spectrum.

PrgmNr 3020 - *ADAM6 (106329183_106736911)x3 gene variant in a Charcot Marie-Tooth patient with upper extremity involvement*

View session detail

Author Block: **V. RodrÃguez Machuca**; UdeG, Guadalajara, Mexico
Disclosure Block: **V. RodrÃguez Machuca:** None.
Introduction. Charcot-Marie-Tooth disease (CMT) encompasses a group of inherited disorders characterized by motor and sensory polyneuropathy. Most of these disorders exhibit an autosomal dominant pattern of inheritance and are classified with respect to neuroconduction studies into 3 major groups. CMT dominant intermediate disease type E (CMTDIE) presents the neurological features of CMT, along with manifestations of chronic kidney disease at an early age (proteinuria, segmental and focal segmental glomerulosclerosis (FSGS)). In addition to the deformity in pelvic limbs, the involvement of the hands is characteristic, observing a 'claw-like' appearance/disposition. **Aim.** To present a patient with CMTDIE and upper extremity involvement and the ADAM6 (14q32.33) (106329183_106736911)x3 gene variant. **Clinical summary.** Male patient, 32 years old. Non consanguineous parents, no family history of malformations or syndromes except for a similarly affected twin brother who declined further examination. Normal psychomotor development until the age of 5 years, with later neuromuscular deterioration. Physical examination revealed upper extremities with decreased strength, hyporeflexia, claw hands, areflexia and abolition of strength in lower limbs. Generalized muscle atrophy was also observed. **Material and methods.** Routine laboratory, neuroconduction studies, nuclear magnetic resonance, karyotype, Short Tandem Repeats (STRs) determination and CGH microarray. **Results.** Urinalysis and Blood chemistry showed a preserved renal function, no evidence of proteinuria. Electromyography NCV Conclusions. Few cases of CMTDIE have been reported. It is infrequent to find claw hands as presented by the patient, although it has been described. In the literature it has been reported that in this variant of CMT there is proteinuria and development of early chronic renal disease, however our patient did not present renal involvement. Noteworthy, according to our review of the literature, there are no reports associating the gene variant presented here with CMTDIE, although the chromosomal region in which this gene is situated has previously been implicated in the disease.

PrgmNr 3021 - *SETBP1 involvement in a neurodevelopmental disorder: A report from a reanalysis of a negative clinical exom*

View session detail

Author Block: **I. Barcelos**[1], D. Li[1], E. McCormick[1], M. J. Falk[2], M. He[3], H. Hakonarson[4]; [1]CHOP, PHILADELPHIA, PA, [2]CHOP, Philadelphia, PA, [3]Childrenâ□□s Hosp. of philadelphia, Philadelphia, PA, [4]Children S Hosp. of Philadelphia, Philadelphia, PA

Disclosure Block: **I. Barcelos:** None.

A range of clinical findings have been associated with heterozygous mutations in the SET-binding protein gene (*SETBP1*). Germline nonsense or truncating variants cause intellectual disability with speech delay, seizures, and mild facial dysmorphisms. Germline recurrent missense variants at residues 868-871, forming the critical consensus sequence of the degradation signal, have been associated with Schinzel-Giedion syndrome, a rare condition characterized by multiple malformations, severe neurological alterations, and increased risk of malignancy. Interestingly, however, somatic recurrent missense variants at residues 868-871 have been shown to cause atypical chronic myeloid leukemia through abrogating ubiquitination, which in turn results in reduced PP2A activity and increased proliferation rates. It was also reported that several affected individuals with germline missense variants outside the degradation-related critical motif exhibit a milder phenotype, including developmental delay, autistic features, and milder facial dysmorphisms. We report an 8 years old male who presented for an evaluation due to developmental delay, autism, hypotonia, hypospadias, and VSD. Previous clinical exome sequencing of this patient was non-diagnostic. We undertook reanalysis of the same WES data in the research setting where we identified a *de novo* missense mutation, c.2572G>A (p.E858K), in *SETBP1*, which was confirmed by Sanger sequencing. This variant was not found in population allele frequency databases, and is predicted to be deleterious by multiple algorithms. Functional assessments of the patient cells highlighted reduced mTORC1 signaling and aberrant autophagy, suggesting a disruption of coordinated action of PP2A and mTORC1 in regulating autophagy. The variant was classified to be pathogenic and this patient shares phenotypic characteristics with previously reported patients *SETBP1* mutations who are mildly affected. It has been well established that reanalysis of exomes may further increase the molecular diagnostic rate. In this report, we adapted our exome analytic pipeline to a patient with nonspecific developmental disorder and established molecular diagnosis, identifying a pathogenic variant in *SETBP1*. Our results demonstrate the benefit of reanalysis of previously obtained sequencing data in a patient who is otherwise lacking a clinical diagnosis.

PrgmNr 3022 - *TBCK*-related encephalopathy: Understanding the role of autophagy in the pathology of the disease

View session detail

Author Block: **S. Murali**[1], R. Angireddy[2], E. Bhoj[2]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **S. Murali:** None.

TBCK-related encephalopathy, also known as *TBCK* syndrome, is a rare pediatric neurodegenerative disorder affecting the central and peripheral nervous systems. It is an autosomal recessive disease caused due to homozygous loss-of-function mutations in the *TBCK* gene. Children affected by this syndrome show phenotypes of intellectual disability and hypotonia. Other symptoms include coarse facial features, global developmental delay (ranging from moderate to severe), difficulty breathing, epilepsy, osteoporosis, and hypothyroidism. TBC1 domain containing kinase (*TBCK*) gene, located on chromosome 4, encodes for the TBCK protein composed of three functional domains: an N-terminal Serine/Threonine kinase domain, a central TBC domain, and a C-terminal rhodanese homology domain (RHOD). The exact function of the TBCK protein is not known, but preliminary studies indicate that the TBC domain functions as a GTPase -activating protein for small Rab GTPases. Studies from patient-derived fibroblast show that the loss of *TBCK* modulates the mechanistic Target of Rapamycin (mTOR) pathway and significant reductions in the downstream effectors of the pathway were observed. *TBCK* patient fibroblast showed an increase in autophagy flux, providing evidence that autophagic-lysosomal dysfunction may play an essential role in the disease's pathophysiology. However, the exact function of this gene in the autophagy process is not known and remains to be explored. The present study is aimed to understand autophagy in *TBCK* neurons. We used human immortalized neural progenitor cell (NPC) line ReNcell VM to stably knockdown *TBCK* using lentivirus transduction of specific shRNA and differentiated NPC cells into dopaminergic neurons. Immunocytochemistry studies will be performed to visualize changes in the formation of autophagosomes and autophagy flux in dopaminergic neurons.

PrgmNr 3023 - A cross-disorder and evidence-based tiered ranking of candidate genes for neuropsychiatric disorders

View session detail

Author Block: **H. Shimelis**[1], M. T. Oetjens[2], M. Znidarsic[1], K. E. Wain[2], S. Myers[1], D. Ledbetter[1], C. L. Martin[2]; [1]Autism & Dev.al Med. Inst., Geisinger, Lewisburg, PA, [2]Geisinger, Lewisburg, PA
Disclosure Block: **H. Shimelis:** None.
Sequencing studies of cohorts with neuropsychiatric disorders (NPD) continue to identify many new NPD genes. These studies have revealed that categorical disorders, including intellectual disability (ID), autism spectrum disorder (ASD), and schizophrenia (SCZ), share genetic etiologies. Several databases have been created to catalogue rare variants in NPD probands, serving as a resource for researchers studying NPD. However, these databases generally use a single categorical NPD diagnosis in their approach, which is inconsistent with the known genetic overlap of these conditions and may underrepresent evidence for a given gene. To address this gap, the Developmental Brain Disorder (DBD) Gene Database (https://dbd.geisingeradmi.org/) uses a cross-disorder and tiered genotype-phenotype data mining approach to identify novel candidate genes and provide further evidence to genes previously implicated in NPD. Here, we present an update of the DBD Gene Database which contains data from 1172 studies published 2003-2020 representing 6481 individuals with pathogenic loss-of-function (pLOF) variants in 649 genes. pLOF and single gene copy number variants are curated from published sequencing studies across six NPD phenotypes: ID, ASD, epilepsy, attention deficit hyperactivity disorder, SCZ, and bipolar disorder. All genes are ranked into four tiers based on the number of cases with *de novo* pLOF variants: Tier 1, the strongest level of evidence, includes genes with three or more *de novo* pLOF variants; Tier 2, genes with two *de novo* pLOF variants; Tier 3, genes with one *de novo* pLOF variant; and Tier 4, genes with only inherited (or unknown inheritance) pLOF variants. Autosomal recessive (AR) genes are curated separately from LOF tier rankings. In our latest update from May 2021, 177 were ranked as Tier 1, 79 as Tier 2, 114 as Tier 3, 147 as Tier 4, and 132 as AR genes. To examine whether genes ranked as Tier 1 were novel or previously recognized as high-confidence genes in other databases, we compared Tier 1 genes to two NPD-related databases: SFARI gene and DDG2P (accessed May 2021). Of the 177 Tier 1 genes, 145 were listed as genes with the highest confidence of being NPD-related in at least one of the two databases while 32 were ranked lower (n=30) or not yet included (n=2). When we evaluated phenotypes of individuals with pLOF variants in Tier 1 genes, 96% (170/177) of genes were associated with 2 or more disorders. These results show that using a cross-disorder approach to NPD gene discovery increases the yield of genes with strong evidence of being NPD-related. Furthermore, our results show evidence for phenotypic heterogeneity in individuals with a pLOF variant in the same gene.

PrgmNr 3024 - A late-onset seizure phenotype in a mouse model of *CHD8* haploinsufficiency

View session detail

Author Block: **M. Krenzer**[1], L. Nguyen[1], C. Weiss[2], B. Bistis[1], M. R. Healey[3], M. Rosales Larios[1], D. Rivas[4], A. Bordey[1], J. P. Noonan[1], R. A. Muhle[5]; [1]Yale Sch. of Med., New Haven, CT, [2]Barnard Coll. of Columbia Univ., New York, NY, [3]Brown Univ., Providence, RI, [4]New York State Psychiatric Inst., New York, NY, [5]Columbia Univ.; New York State Psychiatric Inst., New York, NY
Disclosure Block: **M. Krenzer:** None.
Damaging variants in the chromatin remodeler *Chromodomain Helicase DNA-binding protein 8* (*CHD8*) are highly associated with an increased likelihood for autism spectrum disorder (ASD) diagnosis. People diagnosed with ASD are at risk of epileptic seizures and associated premature mortality. In phenotypic cohorts of people harboring *CHD8*-disruptive mutations, 13-30% reported abnormal electroencephalographic (EEG) activity or seizures, suggesting an etiologic link between damaging *CHD8* mutations and abnormal brain activity.

We developed a mouse model of *CHD8* haploinsufficiency by deleting exon 3 of *CHD8* via CRISPR gene editing. *Chd8*$^{+/-}$ mice express 40-60% of the gene product, have a reduced body weight, and a normal life expectancy. Here we report that more than 30% of *Chd8*$^{+/-}$ mice develop late-onset spontaneous seizures in adulthood during routine handling. These events range from mild staring spells to generalized tonic-clonic seizures, from which there is rapid recovery to baseline. Mice observed to have seizures have a normal lifespan relative to wild-type (WT) littermates and *Chd8*$^{+/-}$ mice that did not have behavioral seizures during handling. Seizures occurred in male and female mice, with male *Chd8*$^{+/-}$ mice having a younger age of onset than female *Chd8*$^{+/-}$ mice.

To assess the window of increased seizure susceptibility, we performed a pharmacological challenge assay with the $GABA_A$ antagonist pentylenetetrazole at 6, 11, and 13 months of age. We found that male *Chd8*$^{+/-}$ mice exhibited a significantly reduced seizure threshold and more severe seizures than WT littermates, with increasing severity in older age. Female *Chd8*$^{+/-}$ mice assessed in the same assay did not show a reduced seizure threshold compared to WT littermates, with seizure severity similar to WT male mice. Video-EEG recordings and EEG quantification at 12 and 18 months revealed a trend towards higher Delta power (1.6-4.0 Hz), reflecting an overall increase in relative slow wave brain activity.

To our knowledge, this is the first report of a late-onset seizure phenotype in a mouse model of *CHD8* haploinsufficiency. Our results show a male sex bias in seizure susceptibility to $GABA_A$ antagonism, pointing to an excitation/inhibition imbalance.

These results contribute to a better understanding of the complex interaction and presumed common mechanisms underlying both ASD and seizures. Our findings highlight the need to monitor people who carry damaging variants in *CHD8* or other genetic risk factors for late developing seizures that may impact health and quality of life.

PrgmNr 3026 - A screening pipeline to identify functional impacts of novel *KCNT1* variants in pediatric epilepsies

View session detail

Author Block: **C. D. Whelan**[1], Z. Zhu[1], T. Danker[2], M. Rizzo[1], V. Nguyen[1], D. R. Bearden[3], J. J. Cherry[1], C. A. Hinckley[1]; [1]Biogen Inc., Cambridge, MA, [2]NMI TT Pharmaservices, Markwiesenstrasse, Reutlingen, Germany, [3]Univ. of Rochester, Rochester, NY
Disclosure Block: **C.D. Whelan:** Salary/Employment; Biogen.
Rationale: Genetic variants at *KCNT1* are associated with a spectrum of epilepsies, including epilepsy of infancy with migrating focal seizures (EIMFS) and sleep-related hypermotor epilepsy (SHE). To date, over 30 distinct pathogenic missense variants at *KCNT1* have been reported in the literature. The majority of variants are hypothesized to exert a pathologic gain-of-function mechanism, increasing KCNT1 (Slack, KNa1.1) channel activity by multiple mechanisms.
Methods: In order to identify novel, potentially pathogenic variants, we analyzed genetic data from pediatric epilepsy patients screened via Invitaeâ□□s Behind the SeizureÂ® Epilepsy Panel (https://www.invitae.com/en/behindtheseizure/). *KCNT1* variants identified via the Invitae panel were subsequently expressed in xenopus oocytes; KCNT1 current-voltage relationships were then quantified with two-electrode voltage clamp recordings using the automated Robocyte system.
Results: We identified 88 unique *KCNT1* variants via the Invitae panel, including 12 patients with six previously reported pathogenic mutations, three patients harboring a novel, likely pathogenic, unreported variant, and 93 patients carrying 81 novel variants of unknown significance (VUS). We subsequently conducted a literature search, identifying 17 additional epilepsy-associated *KCNT1* variants for which in-vitro validation of a gain-of-function mechanism was not previously established. To validate a screening pipeline for uncharacterized, potentially pathogenic mutations, we compared the biophysical properties of novel and/or uncharacterized *KCNT1* variants with those of known, previously characterized pathogenic mutations. Relative to wild type KCNT1, we found increased currents at +70 mV for previously characterized EIMFS variants, while a common *KCNT1* missense single nucleotide polymorphism (T737M) minimally impacted current amplitude. Among previously uncharacterized *KCNT1* mutations, we identified 9 variants with gain-of-function increases in current amplitude.
Conclusions: Our dataset is the largest systematic characterization of *KCNT1* variants to date. These ongoing efforts have identified at least 9 previously uncharacterized *KCNT1* variants as gain-of-function, further supporting the therapeutic rationale for *KCNT1* down-regulation in EIMFS and SHE. Further studies systematically examining pathogenic and non-pathogenic variants will enable a deeper understanding of the structural basis of *KCNT1* gain of function.

PrgmNr 3027 - Autism spectrum disorder: *Magel2* gene variants

View session detail

Author Block: **J. Pascual**[1], O. J. Quincoces[1], S. Carlo[2], C. I. Hernandez[3], A. M. Alejandro[4], E. Albino[5], Y. Mojica[6], A. S. Cornier[7]; [1]Ponce Hlth.Sci. Univ., Ponce, PR, [2]Ponce Hlth.Sci. Univ., Cabo Rojo, PR, [3]Natl. Children's Hosp., Pediatrics Dept., Washington, DC, [4]Dept. of Biology, Univ. of Puerto Rico, Mayaguez Campus, Mayaguez, PR, [5]Univ. of Puerto Rico - Med. Sci. Campus, Caguas, PR, [6]Dept. of Biology, Univ. of Puerto Rico Rio, Piedras Campus, San Juan, PR, [7]San Jorge Children's Hosp. / Ponce Hlth.Sci. Univ., San Juan, PR

Disclosure Block: **J. Pascual:** None.

We are presenting the case of a 17-year-old female of Hispanic descent who was referred to the Genetics clinic when she was 13-year-old due to a diagnosis of autism spectrum disorder. Initial metabolic laboratory workup was negative as well as karyotype, analysis *Fragile X* and chromosomal microarray studies. Autism Disorder-Intellectual disabilities gene panel showed mutations on the *MAGEL2* and *SMARCA4* genes. *MAGEL2* is located in a chromosomal region strongly associated with *Prader-Willi syndrome* (PWS). Mutations in this gene have been associated with *Chitayat-Hall* (CHS) and *Schaaf-Yang* (SYS) syndromes. *SMARCA4* mutations have been associated with *Cofï¬n-Siris Syndrome* (CSS). Both CHS and SYS have been described as part of a phenotypic continuum underlining the possible pleiotropy that mutations in this gene produces. It has also been described that *MAGEL2* complete deletion mutations rather than truncating mutations lead to a milder phenotype. This suggests a possible dominant-negative effect or a leaky expression of the maternal *MAGEL2* allele which has been shown to happen with the *NDN* allele which is also found in the PWS region.

We believe that some clinical features present in this patient are due to the *MAGEL2* mutation. These features include skin picking, anxiety, autism spectrum disorder, intellectual disabilities, neonatal hypotonia, and distal digital hypoplasia. These clinical features are consistent with *MAGEL2* gene mutations described in the medical literature. The loss of heterozygosity in this autosomal dominant gene is responsible for some if not all of the patientâ□□s phenotype. Proteomic studies may provide more information to this genotype-phenotype correlation.

PrgmNr 3028 - Biallelic inframe deletion of human *SOX4* is associated with developmental delay, hypotonia and intellectual disability

View session detail

Author Block: **A. Ghaffar**[1], F. Rasheed[2], M. Rashid[2], H. Van Bokhoven[3], Z. M. Ahmed[1,4], S. Riazuddin[5], S. Riazuddin[1,4]; [1]Otorhinolaryngology Head & Neck Surgery, Sch. of Med., Univ. of Maryland, Baltimore, MD, [2]Natl. Ctr. of Excellence in Molecular Biology, Univ. of the Punjab, Lahore, Pakistan, [3]Dept. of Human Genetics, Donders Inst. for Brain, Cognition and Behavior, Radboud Univ. Med. Ctr., 6500HE, Nijmegen, Netherlands, [4]Dept. of Biochemistry and Molecular Biology, Sch. of Med.,Univ. of Maryland, Baltimore, MD, [5]Jinnah Burn and Reconstructive Surgery Ctr., Allama Iqbal Med. Coll., Univ. of Hlth.Sci., Lahore, Pakistan

Disclosure Block: **A. Ghaffar:** None.

Background: Intellectual disability (ID) is defined as person's limitation in intellectual functioning and adaptive behavior with onset before the age of 18 years, having worldwide prevalence of 1-3%. Among the monogenic causes, autosomal recessive ID (ARID) genes are responsible for more than 50% ID. **Purpose:** The purpose of this study was to identify the genetic causes of early onset ID in a large Pakistani kindred (PKMR225), and to investigate the expression and functional importance of identified ID-associated gene in human brain. **Methods:** Physical examination for ID phenotype was assessed along with MRI and biochemical assays in patients. Whole exome sequencing (WES) was used to identify the candidate pathogenic variant in family PKMR225, further confirmed through sanger sequencing. In-silico analysis and 3-dimensional (3D) molecular modeling were performed to assess the impact of the identified deletion on the encoded protein using tools like; Clustal Omega, MetaDome, ITasser, PyMol. Finally, RNA-based expression analysis in brain was done using GTEx and UCSC single cell browser.**Results:** Both siblings of family PKMR225 have moderate to severe ID with global developmental delay and hypotonia, however they had normal range values for biochemical assay and no brain abnormality was observed in MRI. Through WES, we identified a novel in-frame homozygous deletion variant [c.730_753del24; p.(Ala244_Gly251del)] in *SOX4*, segregating with the phenotype in family PKMR225. SOX4 belongs to group C of the SOX transcription regulating family known to be involved in neurogenesis, skeletogenesis, cardiac and reproductive system development. In-silico analysis show loss of seven predicted hydrogen bonds supporting the probability of structural fold change in an intolerant and conserved region of SOX4. Further, with high overlapping RNA expression analysis of *SOX4* with *SOX11, DCX* and *TRB2* we showed the susceptibility of protein importance in neurogenesis. **Significance:** Our study represents the first report of ARID caused by biallelic pathogenic variant in *SOX4* occurring in non-HMG domain of protein. Previously, four de novo heterozygous variants in *SOX4*, specific to highly conserved structured HMG domain were reported in index cases with ID and mild facial dysmorphism. Thus, this study expands the mutational landscape of *SOX4* and the repertoire of the known genetic causes of ARID.

PrgmNr 3029 - Dysregulation of Canonical and Alternative Replication Protein A Complexes in Huntington Disease and Spinocerebellar ataxia Type 1 brains is associated with CAG instability and phenotype

View session detail

Author Block: **T. Gall-Duncan**[1], J. Luo[2], K. Fujita[3], C-M. Jurkovic[4], V. Li[5], S. Tran[1], H. Tanaka[3], D. Leib[6], S. Lanni[2], N. Sato[7], G. B. Panigrahi[8], T. Prasolava[1], M. Curtis[9], L. Ranum[10], B. L. Davidson[11], E. Campos[1], A. R. La Spada[12], F. Boisvert[4], R. Faull[9], H. Okazawa[3], M. Wold[13], C. E. Pearson[14]; [1]Peter Gilgan Ctr. for Res. and Learning, Toronto, ON, Canada, [2]The Hosp. for Sick Children, Toronto, ON, Canada, [3]Tokyo Med. and Dental Univ., Tokyo, Japan, [4]UniversitÃ© de Sherbrooke, Sherbrooke, QC, Canada, [5]Toronto, ON, Canada, [6]Children's Hosp. of Philadelphia, Philadelphia, PA, [7]Tokyo Med & Dental Univ, Tokyo, Japan, [8]Peter Gilgan Ctr. for Res. and Learning, The Hosp. for Sick Children, Toronto, ON, Canada, [9]Univ. of Aukland, Aukland, New Zealand, [10]Univ Florida, Gainesville, FL, [11]The Children's Hosp. of Philadelphia, Philadelphia, PA, [12]Univ. of California, Irvine, Irvine, CA, [13]Univ. of Iowa, Iowa, IA, [14]Hosp for Sick Children, Toronto, ON, Canada
Disclosure Block: **T. Gall-Duncan:** None.

Expansions of gene-specific CAG/CTG DNA repeats cause >15 neurodegenerative diseases, including Huntington Disease (HD) and spinocerebellar ataxia type 1 (SCA1). Inherited expansions continue to somatically expand as patients age, through a poorly understood mechanism. Larger expansions hasten disease onset and worsen severity and progression, so understanding the molecular processes of expansions is crucial to understanding pathogenesis. Somatic expansions may be regulated by tissue-specific expression of DNA repair proteins. Paradoxically, DNA repair proteins may exacerbate somatic expansions by incorrectly mediating repair of the expansion. We assessed the role of two single-strand DNA binding protein complexes in the expansion process. The canonical single-strand DNA binding complex in humans, replication protein A (RPA), composed of RPA1-RPA2-RPA3, is essential for DNA replication, repair, and recombination. Humans also express a poorly understood primate-specific alternative RPA complex (Alt-RPA) in which RPA4 replaces RPA2. Here we show RPA and Alt-RPA are differentially upregulated in HD and SCA1 patient brain regions, with Alt-RPA demonstrating up to 10-fold upregulation in the most affected brain tissues. In vitro repair of slipped-CAG structures, a DNA intermediate of expansions, shows that high concentrations of RPA enhance repair while high concentrations of Alt-RPA blocks repair. Coincidentally, while both Alt-RPA and RPA bind slipped-CAG structures, RPA efficiently melts slipped-DNAs while Alt-RPA does not. Conducting the first BioID interactomes for the RPA subunits we identified that RPA interacts with proteins known to protect against CAG-associated neurodegeneration, while Alt-RPA interacts with proteins which promote CAG-associated neurodegeneration and CAG expansions, including MSH3. We demonstrate that RPA overexpression completely inhibits somatic repeat expansions in vivo in the striatum of SCA1 mice, coinciding with reductions in biomarkers of CAG disease such as genome-wide DNA damage and mutant Ataxin-1 aggregation in striatal medium spiny neurons. Previously we demonstrated Rpa1 overexpression rescued motor phenotypes and ameliorated elevated DNA damage in cerebellar Purkinje neurons, in the same mice. Our new data demonstrate that somatic expansions in the striatum may contribute to SCA1 mouse phenotypes, and that RPA is an active player in suppressing CAG expansions and pathogenesis, effectively preventing somatic CAG expansions and molecular phenotypes through non-replication mechanisms. In contrast, Alt-RPA likely plays the opposite role by enhancing expansions and pathogenesis.

PrgmNr 3030 - Evaluating digenic inheritance of bi-allelic *PGAP2* and *PGAP3* variants in a case of Mabry syndrome

View session detail

Author Block: **M. D. Thompson**[1], X. Li[2], T. Kinoshita[2], Y. Murakami[2], C. Thomas[3]; [1]Univ of Toronto, Toronto, ON, Canada, [2]Osaka Univ., Osaka, Japan, [3]Yale Univ., TorontNew Haven, CT
Disclosure Block: **M.D. Thompson:** None.
At least six genes are associated with hyperphosphatasia mental retardation syndrome (HPMRS). Disruption of four phosphatidylinositol glycan (PIG) biosynthesis genes, *PIGV*, *PIGO*, *PIGW* and *PIGY*, expressed in the endoplasmic reticulum, result in HPMRS 1, 2, 5 and 6; disruption of the two post GPI attachment to proteins genes, *PGAP2* and *PGAP3*, destabilizes the glycophosphoinositol (GPI) anchored proteins (AP) association with the Golgi membrane, resulting in HPMRS 3 and 4, known as Mabry syndrome, the first of 21 glycophosphoinositol (GPI) disorders (IGD). Here we present a case report of a child with features of HPMRS whose genome underwent targeted exome sequencing to identify *PGAP2* c:284A>G homozygous variant of unknown significance and *PGAP3* c:259G>A homozygous variant of unknown significance. We conducted a rescue assay in *PGAP2* and *PGAP3* deficient CHO cell lines in order to assay the potential pathogenicity of these variants. In the rescue assay with the strong (pME) promoter, the PGAP2 variant almost abolished the activity in CHO cells and the mutant protein could not be detected. The mutation made the *PGAP2* protein unstable. Flow cytometric analysis showed that CD59 and CD55 expression on the *PGAP2* deficient cells transfected with the mutant *PGAP2* construct was not restored at all. We could not detect the decreased activity in *PGAP3* mutant with the strong (pME) promoter driven construct. Using the weaker promoter, the *PGAP3* variant may affect PGAP3 activity mildly. In this case of Mabry syndrome, therefore, the phenotype may be predominantly HPMRS4: resulting from autosomal recessive inheritance of NM_001256240.2 *PGAP2* c:284A>G. We discuss the strategies required to establish evidence of digenic inheritance in GPI deficiency disorders.

PrgmNr 3031 - Fibroblasts and lymphocytes are superior options to whole blood and other peripheral tissues for Mendelian neurological disease diagnosis through RNA-seq

View session detail

Author Block: **S. Chen**[1], M. C. Danzi[1], M. Synofzik[2], R. Schule[3], U. Ulmer[4], S. Zuchner[5]; [1]Dept. of Human Genetics, Univ. of Miami, Miami, FL, [2]Hertie Inst. for Clinical Brain Res., Tuebingen, Baden WÃ¼rttemberg, Germany, [3]Ctr. for Neurology and Hertie Inst. for Clinical Brain Res., TÃ¼bingen, Germany, [4]Hertie Inst. for Clinical Brain Res., Tubingen, Germany, [5]Univ. of Miami, Miami, FL
Disclosure Block: **S. Chen:** None.
Genetic diagnosis of patients with rare Mendelian diseases remains challenging. RNA sequencing (RNA-seq) data from patient cells is often additionally leveraged to aid identification of causal variants. RNA-seq data from the affected neural tissue would be most informative for such analyses, however collecting such samples from living patients is typically not possible. It is imperative to balance ease of collection of patient tissues with their clinical utility. Therefore, we evaluated whole blood, patient-derived fibroblasts and lymphocytes as candidate tissues because they are non-invasive and easy to collect from patients. To identify whether these tissues express genes related to three rare neurological Mendelian diseases (Charcot-Marie-Tooth (CMT), Ataxia and hereditary spastic paraplegia (HSP)) in healthy conditions, we analyzed RNA-seq data from three neural tissues (cortex, cerebellum and tibial nerve) and three peripheral tissues (whole blood, fibroblast, and lymphocyte). This data was accessed from the GTEx sequence repository. Unsurprisingly, neural tissues were found to reliably express a greater percentage of disease-related genes than the peripheral tissues at more than 80%. Encouragingly, fibroblasts and lymphocytes still expressed over 70% of these disease-related genes. Whole blood only had 30% of disease-related genes expressed. To reveal whether fibroblast and lymphocyte tissues can be used to diagnose HSP in a clinical setting, induced pluripotent stem cells (iPSCs), iPSC-derived cortical neurons, fibroblasts and lymphocytes were collected from HSP patients and processed for RNA-seq in addition to genomic sequencing. Gene quantification, splicing detection and differentially expressed gene (DEG) analysis were performed. Based on the splicing and DEG analysis results between four different groups, we identified a list of HSP causal genes that can be reliably detected in fibroblasts and lymphocytes. We also evaluated the potential benefits of using iPSC-derived neurons over fibroblasts and lymphocytes for diagnosis using sets of samples collected from the same patients. Taken together our results indicate that fibroblast and lymphocyte tissue have potential to be non-invasive yet accurate ways to diagnose HSP, CMT and Ataxia.

PrgmNr 3032 - Free Sialic Acid Storage Disease (FSASD) due to defective*SLC17A5*: A status update

View session detail

Author Block: **M. Huizing**[1], M. Hackbarth[2], M. Sabir[3], L. Wolfe[1], M. Wasserstein[4], D. Adams[5], M. V. Malicdan[1], W. A. Gahl[6], S. U. Walkley[7], FSASD Consortium; [1]NIH/NHGRI, Bethesda, MD, [2]NHGRI, ROCKVILLE, MD, [3]NIH NHGRI, Bethesda, MD, [4]Children's Hosp. at Montefiore, Albert Einstein Coll. of Med., Bronx, NY, [5]NIH, Bethesda, MD, [6]NHGRI (NIH), Bethesda, MD, [7]Albert Einstein Col Med, Bronx, NY
Disclosure Block: **M. Huizing:** None.
Free sialic acid storage disorder (FSASD) is an extremely rare, autosomal recessive, neurodegenerative, multisystemic disorder caused by defects in the lysosomal sialic acid (SA) exporter SLC17A5 (sialin). Lysosomal accumulation of free SA in FSASD results in enlarged lysosomes in some cell types and 10-100-fold increased urinary free SA. The clinical spectrum ranges from a severe infantile onset form, lethal in early childhood, to a mild form with subjects living into adulthood, also called Salla disease. In 2018, we initiated a multidisciplinary collaborative effort with clinical and scientific FSASD experts, the National Institutes of Health, and the patient advocacy group (S.T.A.R. Foundation) to overcome scientific, clinical and financial challenges facing the development of treatments for FSASD. Several disease-related facets are now under investigation:
Diagnosis: Around 200 FSASD cases are reported worldwide. FSASD is likely underdiagnosed, due to the rarity of the disorder, absence of routine urine SA testing, and non-specific clinical symptoms. We promote FSASD awareness, *SLC17A5* inclusion in gene panels, and access to SA testing.
Clinical Phenotype: FSASD clinical features include coarse facial features, organomegaly, and progressive neurodegenerative symptoms. Central hypomyelination with cerebellar atrophy and thinning of the corpus callosum are prominent imaging features. To identify diagnostic and clinical trial outcome markers, a FSASD patient registry was initiated and a pilot natural history study is ongoing.
Cellular Phenotype: FSASD pathobiology is poorly understood. To study cellular pathology and functional assays, a variety of live-dye, labeled substrate, and immunofluorescent assays are being tested in patient-derived cells.
FSASD models: Different FSASD cell types are being studied (fibroblasts, neuronal and iPS cells). SLC17A5 knock-out and knock-in mouse models are under investigation.
Therapy: There is no approved therapy for FSASD. We plan to pursue high throughput drug screening, SLC17A5 chaperone testing, and gene editing/therapy approaches.
Encouraged by the success of other collaborative efforts for rare diseases, we aim to accelerate data collection that incentivizes industry to further develop and commercialize FSASD treatments.

PrgmNr 3033 - Genetic evaluation of dementia with Lewy bodies implicates distinct disease subgroups

View session detail

Author Block: **Z. Shah**[1], K. Kaivola[2], R. Chia[1], S. Scholz[1], International LBD Genomics Consortium; [1]NIH, Bethesda, MD, [2]NIH, Bethesda, MD
Disclosure Block: **Z. Shah:** None.
The underlying genetics of dementia with Lewy Bodies (DLB) is not well understood, but studies have shown a strong association between the APOE locus and risk of developing Alzheimerâ□□s disease and DLB. However, there is inconclusive evidence on whether APOE independently drives Î±-synuclein pathology. In this study, we analyzed whole-genome sequence data from 2,466 DLB cases and 2,928 neurologically healthy, aged controls. We first performed an APOE-*Îµ4*-stratified genome-wide association study (GWAS) comparing 1,286 DLB cases without APOE ?4 to 2,271 controls without APOE *Îµ4* and 1,180 DLB cases with APOE ?4 to 657 controls with APOE *Îµ4*. GBA (rs2230288, p = 6.58x10-9, odds ratio [OR] = 3.41, 95% confidence interval [CI] = 2.25-5.17) was the only locus that reached genome-wide significance in the GWAS comparing DLB cases and controls without APOE *Îµ4*. There were no loci that reached genome-wide significance in the GWAS comparing DLB cases and controls with APOE *Îµ4*. Next, we divided 495 DLB cases into three groups based on the severity of the AD co-pathology: pure DLB (n=88), DLB with intermediate AD co-pathology (DLB + iAD, n=66), and DLB with high AD co-pathology (DLB + AD, n=341). For each of the groups, we examined the association of APOE *Îµ4* against the 2,928 neurologically healthy controls. We found that APOE *Îµ4* was strongly associated with DLB + AD (p = 1.29x10-32, OR = 4.25, 95% CI = 3.35-5.39) and DLB + iAD (p = 0.0011, OR = 2.31, 95% CI = 1.40-3.83), but not with pure DLB (p = 0.31, OR = 0.75, 95% CI = 0.43-1.30). Our findings suggest that APOE *Îµ4* is not an independent driver of Î±-synuclein pathology in pure DLB, but rather the association is explained with AD co-pathology. Moreover, GBA is the main risk factor for patients with DLB with little or no AD co-pathology. Future DLB genetic studies should consider the severity of AD co-pathology.

PrgmNr 3034 - Genome wide meta-analysis of suicide behaviors

View session detail

Author Block: **Q. S. Li**[1], A. Docherty[2], A. A. Shabalin[3], E. DiBlasi[4], S. Gopal[1], E. Daly[1], C. Canuso[1], FinnGen, International Suicide Genetics Consortium, W. Drevets[5], H. Coon[4]; [1]Janssen R&D, LLC, Titusville, NJ, [2]VCU, Richmond, VA, [3]Univ. of Utah, Salt Lake City, UT, [4]Univ. of Utah Sch. of Med., Salt Lake City, UT, [5]Janssen R&D, LLC, San Diego, CA

Disclosure Block: **Q.S. Li:** Salary/Employment; Johnson & Johnson.

Following the recently published genomic analysis of suicide death data from a large population-ascertained cohort (Docherty et al 2020), we have genotyped another ~1200 samples from the original cohort and added three cohorts of subjects with a history of suicide attempt as the clinical phenotype (FinnGen cohort and two Janssen clinical trial samples). We aim to identify genetic risk variants associated with suicidal behavior. For the expanded suicide death analysis using 3765 cases and 6572 populational controls genotyped using a matching array platform, and meta-analysis across both suicidal behavior phenotypes (n = 8315 cases and 256,478 psychiatric or populational controls), one locus in neuroligin 1 (*NLGN1*) passing genome wide significance threshold for suicide death phenotype compared to the general population was identified (top SNP rs73182688, with p = 5.48×10^{-8} before and $p = 4.55 \times 10^{-8}$ after mtCOJO analysis conditioning on major depressive disorder (MDD) (Howard et al., summary statistics without the 23andMe cohort). Conditioning on suicide attempt (suicide attempt summary statistics from ISGC excluding suicide death cohorts) did not significantly change the association strength ($p = 6.02 \times 10^{-8}$), suggesting this is a suicide death specific genetic risk locus. *NLGN1* encodes a member of a family of neuronal cell surface proteins. Members of this family may act as splice site-specific ligands for beta-neurexins and are putatively involved in the formation and remodeling of central nervous system synapses. MAGMA gene-set analysis of suicide death suggests enrichment of several immune related pathways such as MHC class Ib receptor activity ($p = 1.82 \times 10^{-7}$, Bonferroni adjusted p-value = 0.003), natural killer cell cytokine production ($p = 7.68 \times 10^{-6}$), positive regulation of interferon gamma secretion, $p = 3.39 \times 10^{-5}$). MAGMA gene-based association tests additionally identified *ROBO2* and *ZNF28* as being associated with suicidal behavior in the meta-analysis across 5 cohorts, among which *ZNF28* was also associated with suicide death. Based on the current analysis, total Liability scale h^2 for suicide death and suicidal attempt are 0.08 and 0.04, respectively. The genetic correlation of suicide death with suicidal attempt and MDD are 0.69 ($p = 1.58 \times 10^{-6}$) and 0.51 (p = 2.92 x 10-6), respectively. Bidirectional Generalised Summary-data-based Mendelian Randomisation (GSMR) analysis suggests that suicidal attempt and suicide death are both bi-directionally causal for MDD. In summary, three genome wide significant findings were identified in this study and immune-related biological pathways have been implicated by gene-set based analysis.

PrgmNr 3035 - Heterozygous Loss-of-Function Variants in TBCK Cause a Mild Neurologic Syndrome in Humans and Mice

View session detail

Author Block: **A. Diaz-Rosado**[1,2], A. Black[1], D. Nair[1], S. Murali[3], E. J. Bhoj[4]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Univ. of Pennsylvania, Phialdelphia, PA, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]Philadelphia, PA

Disclosure Block: **A. Diaz-Rosado:** None.

TBC1 domain-containing Kinase (TBCK)-related encephalopathy, is a rare pediatric neurodegeneration disorder with no current treatments or therapies. Recessive homozygous loss-of-function mutations in this gene (*TBCK*) cause a wide spectrum of pathologies that include brain atrophy, seizures, hypotonia, respiratory failure and intellectual disability. However, although these pathologies are usually seen in homozygous patients, we here describe, for the first time, a milder undescribed phenotype that is seen in human TBCK LOF heterozygotes (TBCK-/+). Out of family testimonials, our group also found that in some of these individuals, the processing and execution of motoric functions was also disrupted (*apraxia*). Therefore, to test the possibility of a behavioral and molecular phenotype, we decided to employ our heterozygous mouse model (C57B/6J; TBCK-/+) to address this question. Both of our behavioral and molecular data suggest the presence of a mild neurologic syndrome that affects both the behavior and motor output of animals, while in humans, these results, concomitantly with our molecular data, have also revealed a metabolic phenotype that affects the levels of different markers like pAKT and pS6, and TBCK itself. Altogether, these results pinpoint the presence of a mild syndrome that not only has an impact upon heterozygous humans, but that can also have repercussions upon the motor dynamism of mice.

PrgmNr 3036 - Identification of a Novel *CNTNAP2*Variant in a Consanguineous Pakistani Family with Epilepsy

View session detail

Author Block: **K. Mattison**[1], N. Badshah[2], P. Chopra[3], H. R. Johnston[4], S. Ahmad[2], M. E. Zwick[4], A. P. Escayg[4]; [1]Dept. of Human Genetics, Emory Univ Sch. of Med., Atlanta, GA, [2]Inst. of Biotechnology and Genetic Engineering, Univ. of Agriculture Peshawar, Peshawar, Pakistan, [3]Emory Univ, Atlanta, GA, [4]Emory Univ Sch. of Med., Atlanta, GA

Disclosure Block: **K. Mattison:** None.

We performed whole-genome sequencing on two Pakistani brothers from a consanguineous pedigree of Afridi tribal heritage. The two brothers presented with epilepsy, intellectual disability, and autism spectrum disorder. Through WGS we identified approximately 80,000 variants within shared regions of homozygosity between the two brothers. Of these 80,000 variants, 615 had a minor allele frequency â¤ 0.001 and 21 of those variants had CADD scores â¥ 15. Four homozygous, exonic variants were identified in both affected brothers: *PDZD7* (c.1348_1350delGAG, p.Glu450del), *ALG6* (c.1033G>C, p.Glu345Gln), *RBM20* (c.1587C>G, p.Ser529Arg), and *CNTNAP2* (c.785G>A, p.Gly228Arg). We performed Sanger sequencing of these variants in the parents and two unaffected siblings. The *ALG6* variant was homozygous in one unaffected family member, indicating that it is unlikely to pathogenic. The *PDZD7, RBM20,* and *CNTNAP2* variants were found to be homozygous only in the affected brothers, consistent with the expected recessive mode of inheritance. Pathogenic variants in *PDZD7* and *RBM20* are associated with autosomal recessive non-syndromic hearing loss and autosomal dominant dilated cardiomyopathy, respectively, suggesting that these variants are unlikely to contribute to the neurological clinical presentation in the affected brothers. The clinical presentation of the affected brothers is consistent with that of previously reported patients with homozygous *CNTNAP2* variants, suggesting that the p.Gly228Arg change in *CNTNAP2* is likely causative. This is the first study on the genetic etiology of epilepsy within the Afridi tribe of Pakistan.

PrgmNr 3037 - Impaired *SNAPC4*function leads to global reduction of canonical splicing events and is associated with a disorder characterized by progressive spasticity, developmental delay, and speech dysarthria

View session detail

Author Block: **F. G. Frost**[1], P. Sharma[2], M. Morimoto[2], N. Belnap[3], D. Calame[4], V. Narayanan[5], S. Rangasamy[6], M. J. Huentelman[7], K. Ramsey[7], L. T. Emrick[8], J. R. Lupski[9], P. Steinbach[10], Undiagnosed Diseases Network, J. Murphy[10], E. Macnamara[11], C. J. Tifft[12], W. A. Gahl[13], M. Malicdan[10]; [1]NIH Undiagnosed Diseases Program, Bethesda, MD, [2]NIH, Rockville, MD, [3]The Translational Genomics Res. Inst., Phoenix, AZ, [4]Houston, TX, [5]Translational Genomics Res. Inst., Phoenix, AZ, [6]Translational Genomics Res. Inst. (TGen), Phoenix, AZ, [7]TGen, Phoenix, AZ, [8]Baylor Coll. Med., Houston, TX, [9]Baylor Col Med, Houston, TX, [10]NIH, Bethesda, MD, [11]NIH, Bethesda, MD, [12]Office of the Clinical Director, Bethesda, MD, [13]NHGRI (NIH), Bethesda, MD

Disclosure Block: **F.G. Frost:** None.

Introduction: Small nuclear RNAs (snRNAs) play integral roles in RNA splicing and are critical for cell function. Transcription of snRNAs is initiated by the small nuclear RNA activating protein complex (SNAPc), which binds to a promoter element upstream of snRNA genes and recruits transcriptional machinery. Zebrafish models that impair transcription of SNAPc subunits led to severe developmental defects or even embryonic lethality, demonstrating the critical role of the complex in development. Here we present 4 individuals from 3 families who presented with progressive spasticity, developmental delay, and dysarthria, each with bi-allelic deleterious variants in *SNAPC4*, encoding the SNAPc subunit.

Methods: Detailed clinical and biochemical phenotyping was performed on all probands. Exome and genome sequencing of the probands and their families was performed to identify candidate variants, which were validated by Sanger sequencing. Computational modelling was used to assess the spatial location of variants within the DNA-binding domain. The functional impact of *SNAPC4* variants were assessed using patient-derived fibroblasts and a HeLa *SNAPC4* deficient cell line, generated using CRISPR-Cas9 technology. SNAPc function was assessed by measuring relative snRNA expression with RT-qPCR, and global splicing function was evaluated using RNA-sequencing.**Results:** Four individuals with progressive spasticity, developmental delay, and dysarthria harbored bi-allelic, predicted pathogenic *SNAPC4* variants that segregated with disease. Two variants were in canonical splice sites, that were shown to introduce premature stop codons by cDNA sequencing of patient-derived fibroblasts. Three of the four missense variants are located in the DNA-binding domain of *SNAPC4*, demonstrated by protein modeling to be located on surface residues. Patient fibroblasts showed decreased expression of both *SNAPC4* mRNA and SNAPC4 protein, as well as altered expression of the *RNU1*, *RNU2*, *RNU4*, *RNU5*, *RNU6*, *RNU12*, and *RN7SK* snRNAs. Deficiency of *SNAPC4* in HeLa cells led to a broad decrease in snRNA expression and a decrease in the number of canonical alternative splicing events.**Conclusion**: Our work supports the role of *SNAPC4* in snRNA transcription and splicing in mammalian cells. The common clinical phenotypes and the alteration of *SNAPC4* and snRNA expression in patient cells support mutations in *SNAPC4* as a plausible explanation for disease in our cohort. Identification of additional affected individuals with *SNAPC4* mutations will be required to confirm the genotype-phenotype connection and future experiments will elucidate the underlying mechanisms of disease.

PrgmNr 3038 - Neurofilament light chain in cerebrospinal fluid as a potential novel biomarker in evaluating both clinical severity and therapeutic response in Niemann-Pick Disease, type C1

View session detail

Author Block: **N. Agrawal**[1], N. Farhat[2], A. Dang Do[2], C. Xiao[1], E. M. Berry-Kravis[3], N. Sinaii[4], R. Masvekar[5], B. Bielekova[5], F. D. Porter[6]; [1]NHGRI, NIH, Bethesda, MD, [2]NICHD, NIH, Bethesda, MD, [3]Rush Univ Med Ctr, Chicago, IL, [4]BCES, CC, NIH, Bethesda, MD, [5]NIAID, NIH, Bethesda, MD, [6]NIH, Bethesda, MD

Disclosure Block: **N. Agrawal:** None.

Niemann-Pick Disease, type C (NPC) is caused by recessive pathogenic variants in either NPC1 or NPC2 that lead to intracellular accumulation of glycosphingolipids and cholesterol. Patients with NPC can present with liver disease as neonates or with neurodegeneration at any age. There are currently no biomarkers that correlate well with disease severity and progression, thus monitoring is done using a clinical severity scale [NPC Neurological Severity Score (NSS)]. Neurofilament light chain (NfL) is a structural protein that exists only in neuronal axons and is currently being investigated in a number of neurodegenerative diseases as a potential biomarker, as it is released into cerebrospinal fluid (CSF) as a consequence of neuronal damage. In this study, CSF NfL levels were measured in 116 NPC1 participants. Both cross-sectional and longitudinal data were obtained. Samples obtained from 30 pediatric patients without NPC1 were used as controls. We found that the age-adjusted level of CSF NfL was significantly elevated at baseline (1198.49 Â±1553.11 pg/ml) compared to controls (374.09 Â±633.13; ps=0.28, 95% CI 0.10-0.44, p=0.003). When looking individually at five major domains included in the NSS, worsening symptoms in four of five domains (ambulation, fine motor, speech, swallowing) specifically correlated with increased CSF NfL levels. There was no correlation of CSF NfL levels with the cognition domain. Many of the patients in the study were also being treated off-label with miglustat or with the investigational drug 2-hydroxypropyl-Î²-cyclodextrin (HPÎ²CD, VTS-270). No correlation was observed between CSF NfL levels and treatment with intrathecal HPÎ²CD. In contrast, miglustat therapy was associated with significantly decreased CSF NfL levels (p

PrgmNr 3039 - Pre-clinical studies in induced Pluripotent Stem Cell (iPSC) lines with *SORD* mutations linked to a recessive neuropathy

View session detail

Author Block: **C. Yanick**[1], R. Maciel[1], J. Medina[1], A. P. Rebelo[2], M. Shy[3], S. Zuchner[2], M. Saporta[1]; [1]Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Univ Miami, Miami, FL, [3]Univ. of Iowa, Iowa City, IA
Disclosure Block: **C. Yanick:** None.
Introduction: Recently, mutations in the gene coding for sorbitol dehydrogenase (*SORD*) were associated with a new form of recessive inherited neuropathy. To model this CMT type, we took fibroblast samples from patients and reprogrammed them into induced-Pluripotent Stem Cells (iPSCs). We are using these cells to perform assays to evaluate different therapeutic strategies for SORD related neuropathy. **Methods:** Fibroblasts were transduced using a Sendai Virus Reprogramming kit to express reprogramming factors. Immunocytochemistry for Oct4, NANOG, and SSEA3 was performed to confirm pluripotency of cells. RNA was tested to confirm exogenous genes were no longer present and karyotyping was performed on the iPSC lines. Intracellular sorbitol levels were measured via colorimetric sorbitol kit. Neurofilament Light Chain was measured via ELISA kit. RNA was sequenced by the UM Genomics Core lab. **Results:** Immunocytochemistry (ICC), RNA analysis, Sanger sequencing and karyotyping results showed that we have successfully generated iPSCs from fibroblasts for each of the 3 patients with biallelic *SORD* mutations. We were then able to differentiate these cells into motor neurons in 2D culture to carry out various characterization studies and therapy screenings. We were able to show that the SORD motor neurons show an increase in intracellular sorbitol content when compared to controls, replicating an important disease phenotype. The sorbitol content of patient motor neurons were roughly 2-fold higher than that of patient fibroblasts. Initial measurements of supernatant NFL levels show no increase between control and patient motor neurons. We are currently studying the impact of metabolic stress on the SORD motor neurons, their susceptibility to neurodegeneration and potential transcriptomic dysregulation. Our future plans include performing rescue experiments with various therapeutic strategies using both 2D and 3D cultures. These ongoing efforts include genetic therapies such as ASOs as well as using small molecules to target pathways involving sorbitol metabolism. **Conclusions:** Our results show that we were able to reprogram patient fibroblasts into iPSC lines and then differentiate those lines into motor neurons, that replicate disease phenotype, for our ongoing studies. The ongoing pre-clinical studies of small molecule and genetic therapies will demonstrate target engagement, biomarker availability and lead to translational applications in clinical trials.

PrgmNr 3040 - Truncating variants in the *SHANK1* gene are associated with a spectrum of neurodevelopmental disorders

View session detail

Author Block: **H. J. May**[1], J. Jeong[2], A. Revah Politi[3], J. S. Cohen[4], A. Chassevent[5], J. Baptista[6], E. H. Baugh[1], L. Bier[1], A. Bottani[7], M. Carminho A. Rodrigues[1], C. Conlon[5], J. Fluss[8], M. Guipponi[9], C. A. Kim[10], N. Matsumoto[11], R. Person[12], M. Primiano[1], J. Rankin[6], M. Shinawi[13], C. L. Smith-Hicks[14], A. Telegrafi[12], S. Toy[15], Y. Uchiyama[16], V. S. Aggarwal[17], D. B. Goldstein[18], K. Roche[2], K. Anyane-Yeboa[19]; [1]Columbia Univ. Irving Med. Ctr., New York, NY, [2]NIH, Bethesda, MD, [3]Inst. for Genomic Med.- Columbia Univ., New York, NY, [4]Kennedy Krieger Inst, Baltimore, MD, [5]Kennedy Krieger Inst., Baltimore, MD, [6]Royal Devon and Exeter NHS Fndn. Trust, Exeter, United Kingdom, [7]Geneva Univ Hosp, Geneva 4, Switzerland, [8]Geneva Children's Hosp., Geneva, Switzerland, [9]Univ Geneva Med. Sch., Geneva, Switzerland, [10]Inst da Crianca, Sao Paulo, Brazil, [11]Yokohama City Univ Grad Sch Med, Yokohama, Kanagawa, Japan, [12]GeneDx, Gaithersburg, MD, [13]Washington Univ. Sch. of Med., St Louis, MO, [14]Kennedy Krieger Inst./Johns Hopkins Univ SOM, Baltimore, MD, [15]Washington Univ. Sch. of Med., St. Louis, MO, [16]Yokohama City Univ., Yokohama, Kanagawa, Japan, [17]Columbia Univ, New York, NY, [18]Columbia Univ Med Ctr, New York, NY, [19]Columbia Univ. Med. Ctr., New York, NY

Disclosure Block: **H.J. May:** None.

Purpose: In this study, we aimed to characterize the clinical phenotype of a *SHANK1*-related disorder and define the functional consequences of *SHANK1* truncating variants.

Methods: Exome sequencing (ES) was performed for six individuals who presented with neurodevelopmental disorders. Individuals were ascertained with the use of GeneMatcher and Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (DECIPHER). We evaluated potential nonsense mediated decay (NMD) of two variants by making knock-in cell lines of endogenous truncated SHANK1, and expressed the truncated *SHANK1* cDNA in HEK293 cells and cultured hippocampal neurons to examine the proteins.

Results: ES detected *de novo* truncating variants in *SHANK1* in six individuals. Evaluation of NMD resulted in stable transcripts, and the truncated SHANK1 completely lost binding with Homer1, a linker protein that binds to the C-terminus of SHANK1. These variants may disrupt protein-protein networks in dendritic spines. Dispersed localization of the truncated SHANK1 variants within the spine and dendritic shaft was also observed when expressed in neurons, indicating impaired synaptic localization of truncated SHANK1.

Conclusion: This report expands the clinical spectrum of individuals with truncating *SHANK1* variants and describes the impact these variants may have on the pathophysiology of neurodevelopmental disorders.

PrgmNr 3041 - Using*Drosophila melanogaster*to identify loci modifying Coffin-Siris syndrome mutations

View session detail

Author Block: **R. A. MacPherson**, T. F. C. Mackay, R. R. H. Anholt; Clemson Univ., Greenwood, SC
Disclosure Block: **R.A. MacPherson:** None.
Coffin-Siris syndrome (CSS) and Nicolaides-Baraitser syndrome (NCBRS) are rare disorders of chromatin modification associated with alterations of subunits within the highly conserved mammalian SWI/SNF complex. CSS and NCBRS patients typically present with intellectual disability, facial and digit abnormalities, seizures, and hypotonia. However, phenotypic presentation and disease severity varies within and across CSS- and NCBRS-associated mutations; individuals with identical genetic alterations do not necessarily present with identical phenotypes. We hypothesize that there are naturally occurring genetic variants segregating in the human population that serve as modifiers of disease severity. Identification of candidate genetic modifiers may lead to insights on the pathogenesis and treatment of CSS, NCBRS, and related disorders, but genome wide association analyses that are used to map genetic variants associated with common human diseases require large sample sizes that are not possible for rare diseases. Using a bipartite *UAS-GAL4* RNA interference system in the model organism *Drosophila melanogaster*, we have developed a CSS/NCBRS fly model and show that flies with reduced expression of CSS-associated fly orthologs exhibit changes in sleep, activity, and sensorimotor integration. We have identified genes co-regulated with CSS-associated fly orthologs that exhibit similar changes in behavior to our CSS fly models, suggesting that these co-regulated genes may serve as plausible candidate genetic modifiers for CSS and NCBRS. We are currently working to identify epistatic genetic interactions between CSS-associated fly orthologs and co-regulated genes.

PrgmNr 3042 - Biochemical consequence of deficient glycan extension in ALG3-CDG

View session detail

Author Block: **E. Paul Daniel**[1], A. Edmondson[2], Y. Argon[3], H. Alsharhan[4,5], C. T. Lam[6], K. Raymond[7], H. H. Freeze[8], M. He[9]; [1]Dept. of Pathology and Lab. Med., The Children's Hosp. of Philadelphia,, Philadelphia, Philadelphia, PA, [2]Dept. of Pediatrics, Div. of Human Genetics, The Children's Hosp. of Philadelphia,, Philadelphia, PA, [3]Dept. of Pathology and Lab. Med., The Children's Hosp. of Philadelphia,, Philadelphia, PA, [4]Children's Hosp. of Philadelphia, Ellicott City, MD, [5]Dept. of Pediatrics, Faculty of Med., Kuwait Univ., Kuwait, Kuwait, [6]Seattle, [7]Mayo Med. Lab., Rochester, MN, [8]Sanford Burnham Prebys Med. Discovery Inst., La Jolla, CA, [9]Childrenâ□□s Hosp. of philadelphia, Philadelphia, PA

Disclosure Block: **E. Paul Daniel:** None.

ALG3-CDG is a rare Congenital Disorder of Glycosylation (CDG) with clinical phenotype that includes severe neurological manifestations, immune deficiency and skeletal anomalies. The ALG3 enzyme is required to form lipid liked oligosaccharide (LLO)-Man6GlcNAc2 (Man6) from Man5. N-linked glycosylation is critical in aiding protein folding. Two major N-glycan structures, Man5 and Man9 can be transferred en bloc from LLO to nascent proteins, and the extension from LLO-Man5 to LLO-Man9 is the first and fastest Unfolded Protein Response (UPR) upon endoplasmic reticulum (ER) stress. Following glycan extension, UPR directs three pathway specific transducers: IRE1Î±, PERK and ATP6, which regulate downstream translation repression and activation of ER chaperones such as BiP (GRP78) and GRP94, to aide protein folding. The misfolded proteins bearing Man5 and Man9 are reported to be effectively degraded by ERAD mechanism.

We have used ALG3-CDG patientsâ□□ fibroblasts to measure the expression of UPR and ERAD markers by RT-PCR. We observed 1 to 2-fold increase in basal XBP1 splicing, along with 1-2 fold increase of the gene expression of downstream UPR reporters in ALG3-CDG cells, including GRP94, GRP78, OS9 and EDEM1, consistent with constitutional/chronic activation of UPR. When ER stress was induced by 1 mM DTT, 2-3 fold increase in GRP94 and BiP expression was detected compared with controls. Further, by ddVenus assay that uses an engineered misfolded fluorescent glycoprotein as a reporter, we found that ERAD activity was mildly increased in these cells. We show that N-linked Man3-4 released from cellular glycoproteins, are markedly increased while Man5 is not increased despite the accumulation of LLO-Man5 in ALG3-CDG cells. Similarly, the abundance of Man0-4 GlcNAc2 (Man0-4) is increased in ALG3-CDG plasma samples (n=10). In particular, increased N-linked Man3-4 was detected on purified plasma IgG, suggesting an association between altered glycosylation and chronic UPR and immune deficiency in these patients.

In summary, we provide evidence that there is constitutional activation of UPR in ALG3-CDG cells through IRE1Î± pathway. When ER stress is induced in vitro, both UPR and ERAD increase in these cells. N-linked Man3-4 are increased on both cellular proteins and secreted proteins in plasma, which can serve as important ALG3-CDG biomarkers.

PrgmNr 3043 - Bone cell functions in *PPIB* knock-out mouse model for type IX osteogenesis imperfecta are distinct from classical dominant OI

View session detail

Author Block: **Y. Liu**[1], T. Hefferan[2], N. Fratzl-Zelman[3], J. Marini[1]; [1]Section on Heritable Disorders of Bone & Extracellular Matrix, Natl. Inst. of Child Hlth.and Human Dev., NIH, Bethesda, MD, [2]Biomaterials and Histomorphometry Core Lab., Mayo Clinic, Rochester, MN, [3]Ludwig Boltzmann Inst. of Osteology at the Hanusch Hosp. of OEGK and AUVA Trauma Ctr. Meidling, 1st Med. Dept., Hanusch Hosp., Vienna, Austria
Disclosure Block: **Y. Liu:** None.
Osteogenesis imperfecta (OI) is a genetically heterogeneous bone fragility disorder. While most OI cases are caused by dominant mutations in type I collagen, other cases are caused by recessive defects in genes encoding collagen-interacting proteins. Cyclophilin B (CyPB), encoded by *Ppib*, is an ER-localized peptidyl-prolyl cis-trans isomerase (PPIase). It functions independently as the major PPIase catalyzing collagen folding. CyPB is also a component of the procollagen prolyl 3-hydroxylation (P3H1/CRTAP/CyPB). Mutations in *Ppib* cause recessive type IX OI, with a lethal to severe phenotype. We reported previously that *Ppib*$^{-/-}$ mice have abnormal type I collagen post-translational modification and crosslinks. Unlike other causal genes for OI, *Ppib* is expressed in both bone-forming osteoblasts and bone-resorbing osteoclasts. This study focuses on *Ppib*$^{-/-}$ bone cell functions, utilizing histomorphometry and qBEI analysis of 2-month male femoral tissue, RT-PCR and alizarin red staining of differentiated calvarial osteoblasts, and in vitro osteoclast differentiation.
Histomorphometry of femora from *Ppib*$^{-/-}$ mice reveals markedly reduced cortical bone width (pPpib$^{-/-}$ 5.8Â±1.9%, WT 10.4Â±2.7%; pPpib$^{-/-}$ osteoblast number and surface are significantly decreased (pPpib$^{-/-}$ KO mice have reduced osteoid volume (69%, pn vitro osteoblast mineral deposition is increased (Alizarin red staining, pSost, Mepe, Phex, Dmp1, vs WT. Furthermore, qBEI analysis of bone tissue yielded bone mineralization density distribution with increased CaMean, CaPeak (both pin vitro osteoclast differentiation.
Cyclophilin B/*Ppib* KO mice, modelling type IX OI, have bone cell functions distinct from classical dominant OI with collagen defects. *Ppib* KO bone has a low turnover cellular pattern with decreased osteoblast number and bone formation, increased mineralization, and normal osteoclast numbers.

PrgmNr 3044 - Distribution and frequency of the variant *rs10738445* of *BNC2* gene associated to Idiopathic Scoliosis in Puerto Rico

View session detail

Author Block: **N. GarcÃa TubÃ©ns**, E. G. RamÃrez Aponte, J. C. MartÃnez Cruzado; Univ. of Puerto Rico- MayagÃ¼ez Campus, MayagÃ¼ez, Puerto Rico
Disclosure Block: **N. GarcÃa TubÃ©ns:** None.
Idiopathic Scoliosis is a deformity of the spine, due to the growth of the person, shaped like an â□□Câ□□ or â□□Sâ□□; at a minimum Cobb angle of 10Â°. Recent studies have found a susceptibility allele rs10738445 increases the expression of the BNC2 gene especially in homozygous for the haplotype CC. The BNC2 protein is present in the myoblasts and uterus, spinal cord, bone, and cartilage tissue. Thus, suggesting that there is a relationship between the increase in the gene and the etiology of Idiopathic Scoliosis. Other findings suggest a functional role for BNC2 in the development and progression of spinal deformity in patients with Idiopathic Scoliosis. The SNP rs10738445 of the BNC2 gene appeared in the genomic study 1000 Genomes Project, and Puerto Rico showed a genotypic frequency of 0.53 for the haplotype CC. Using rt-PCR, 622 samples of the general population of Puerto Rico were genotyped for the SNP rs10738445. In this study, we obtained that the genotype CC of rs10738445 has a genotipic frequency of 0.059 in Puerto Rico, with a higher prevalence in the western area. The differences in genotype frequencies could be the result of the bias in the sampling of the Puerto Ricans that participated in the 1000 Genomes Project. We are planning to genotype patients diagnosed with Idiopathic Scoliosis and other spinal deformities for this variant in Puerto Rico.

PrgmNr 3045 - GPC6 haploinsufficiency in patients with skeletal anomalies

View session detail

Author Block: **N. Vasli**[1], A. EDWARD[1], M. J. Nowaczyk[2], C. Li[3], M. KOZENKO[4], D. J. Stavropoulos[5], E. McCready[2]; [1]Dept. of Pathology and Molecular Med., McMaster Univ., Hamilton, ON, Canada, [2]McMaster Univ, Hamilton, ON, Canada, [3]McMaster Children's Hosp, Hamilton, ON, Canada, [4]Dept. of Pediatrics, McMaster Univ., Hamilton, ON, Canada, [5]Hosp for Sick Children, Toronto, ON, Canada
Disclosure Block: **N. Vasli:** None.
Introduction Glypicans are a family of proteoglycans that are covalently bound to the plasma membrane by a glycosylphosphatidylinositol (GPI)-anchor. The mammalian genome consists of six glypicans (*GPC1 - GPC6*). Mutational analysis studies have revealed the role of GPCs in some congenital malformations involving bone growth and heart pattern formation (Fico and Dono, 2005). Biallelic loss-of-function mutations in *GPC6* which altered growth factor signaling and morphogen gradients leading to long bone growth retardation have been reported in patients with autosomal recessive omodysplasia (OMOD1) which is characterized by short stature, craniofacial dysmorphism, and variable developmental delay (Campos-Xavier et al., 2009). **Methods** In this study, a cohort of patients with heterozygous 13q31.3 microdeletions were examined. When DNA was available, compound heterozygosity for *GPC6* pathogenic variants was ruled out by Sanger sequencing of the *GPC6* exonic regions. The clinical indications for chromosome microarray testing were examined in each cohort patient. When available, clinical chart reviews were also performed. Associated clinical features from cohort patients were compared to those reported from previously published cases. **Results** The patients in this cohort had overlapping phenotypes. While skeletal anomalies such as short stature, scoliosis, and short fingers were common among the examined individuals, dysmorphic facies, neurodevelopmental and cardiac abnormalities were also apparent in several patients. Sanger sequencing did not identify pathogenic *GPC6* variants on the retained allele, suggesting that these clinical features are not the result of autosomal recessive omodysplasia. **Discussion** Despite the characterization of autosomal recessive *GPC6* mutations causing omodysplasia, our findings suggest that heterozygous *GPC6* deletions may also be associated with an increased risk of clinical phenotypes, albeit milder than those reported in patients with biallelic variants. These results suggest *GPC6* haploinsufficiency may represent a novel syndrome, characterized by short stature, skeletal anomalies, and variable neurodevelopmental and cardiac abnormalities.

PrgmNr 3046 - Molecular test of Paget's disease of bone in families not linked to *SQSTM1* gene mutations

View session detail

Author Block: **L. Michou**[1,2,3], Y. You[1], D. Simonyan[1], A. Bureau[4], E. Gagnon[1], C. Albert[5], J. R. Guertin[1,6], J-E. Tarride[7], J. P. Brown[1,2]; [1]CHU de QuÃ©bec-UniversitÃ© Laval research centre, Quebec, QC, Canada, [2]Dept. of Med., UniversitÃ© Laval, Quebec, QC, Canada, [3]Dept. of Rheumatology, CHU de QuÃ©bec, Quebec, QC, Canada, [4]Univ Laval, QuÃ©bec, QC, Canada, [5]CHU de MontrÃ©al, Dept. of biochemistry, MontrÃ©al, QC, Canada, [6]DÃ©partement de mÃ©decine sociale et prÃ©ventive, UniversitÃ© Laval, Quebec, QC, Canada, [7]Dept. of Hlth.Res. Methods, Evidence and Impact, Faculty of Hlth.Sci., McMaster Univ., Hamilton, ON, Canada

Disclosure Block: **L. Michou:** None.

Purpose: Paget's disease of bone (PDB) is a focal metabolic bone disorder characterized by an increased bone remodeling. Fifteen to 40% of PDB patients have a familial form with an autosomal dominant inheritance. Disease-causing mutations of the *SQSTM1* gene have been linked to PDB in about 40% of families whereas genes linked to the remaining families are unknown. Several single nucleotide polymorphisms (SNPs) have been associated with PDB in unrelated patient non-carriers of a *SQSTM1* mutation. The current clinical practice guidelines still recommend to measure total serum alkaline phosphatase (sALP) for PDB screening. In unrelated individual non-carriers of *SQSTM1* mutations, we previously developed a genetic test combining male sex with five genetic markers (rs499345, rs5742915, rs2458413, rs3018362, rs2234968), giving rise to an area under the curve (AUC) for PDB phenotype of 0.73 [0.69; 0.77]. A combination of male sex with total calcium corrected for albumin and Procollagen type I N-terminal propeptide (P1NP), had an AUC of 0.82 [0.73; 0.92]. Combining both genetic and biochemical tests increased the AUC to 0.89 [0.83; 0.95]. Objective: This study aimed at testing these genetic and/or biochemical tests of PDB, in families not linked to *SQSTM1* mutations with disease-causing genes yet unknown. Methods: We genotyped the 5 SNPs cited above and measured calcium corrected for albumin, P1NP in 181 relatives, with PDB or not, from 19 PDB families not linked to *SQSTM1* mutations. Bivariate and multivariate logistic regression models adjusted for sex were fitted to search for a molecular test that could best detect PDB in these families. A classification table was generated to establish a cut-off point for continuous variables. Results: Logistic regression estimates of our previous molecular test gave rise to a high sensitivity of 78%, 97% and 88% for the genetic, biochemical and combined test but the specificity was very low, 35%, 11% and 21%, respectively. We then generated in these families new logistic regression estimates but on the same parameters as mentioned above, for the genetic test giving rise to an AUC of 0.65 [0.55; 0.75], for the biochemical test with an AUC of 0.84 [0.74; 0.94], and for the combination test with an AUC of 0.89 [0.82; 0.96], the latter having a sensitivity of 96% and specificity of 57%. The best predictor was the P1NP. Conclusion: In PDB families not linked to *SQSTM1* mutations, the estimates of our previous molecular test gave rise to a poor specificity. Using new estimates, our genetic test was less likely to predict PDB in these families than our previous test but the biochemical and combined tests have similar predictive abilities than our former test.

PrgmNr 3047 - Neonatal presentation of Methylmalonic Acidemia (MMA) with only a single heterozygous mutation in MUT

View session detail

Author Block: **M. Racobaldo**[1], P. Morelli[2], T. Wiseman[1], C. Gordillo-Casero[1], K. Spence[1], A. Sanchez-Valle[3]; [1]Univ. of South Florida, Tampa, FL, [2]Florida State Univ. Coll. of Med., Tallahassee, FL, [3]Univeristy of South Florida, Tampa, FL

Disclosure Block: **M. Racobaldo:** None.

We present a female patient with a single missense mutation in the MUT gene involved in the pathogenesis of Methylmalonic Acidemia (MMA). This report will highlight the importance of clinical diagnosis in the absence of underlying molecular diagnosis and genetic testing in preconception and family planning. MMA is an inborn error of metabolism caused by a deficiency or absence of the enzyme methylmalonyl-CoA mutase (MCM), which is dependent on cobalamin. MCM is required in the breakdown of certain amino acids as well as odd chain fatty acids. MMA belongs to a group of conditions called organic acid disorders where there is a build-up of organic acids in the body that can be toxic when not treated. The effects of MMA usually appear in early infancy and can include vomiting, dehydration, hypotonia, developmental delay, lethargy, hepatomegaly, and failure to thrive. Long-term complications can include feeding problems, intellectual disability, chronic kidney disease, and pancreatitis.Our patient is a 10 months old female who presented to the Emergency Department at day of life 3 for tachypnea and decreased intake and was admitted to the Neonatal Intensive Care Unit for hypovolemic shock and severe metabolic acidosis. At 34 hours of life, her serum ammonia level was 932.5 umol/L (reference range T, p.Gly717Val missense variant results in a loss of function of the enzyme MCM. This mutation is associated with subtype mut(-), indicating partial loss of enzymatic function and B12 responsiveness; This patient has thus continued hydroxycobalamin injections. At her at clinic visit 8 months old, her weight was in the 35th percentile and methylmalonic acid level was 89,600 nmol/L (normal 87-318 nmol/L). She has developmental delays. MMA is an autosomal recessive disease known to be caused by homozygous or compound heterozygous mutations in the MUT gene. Only one mutation was found in this patient; It is assumed that another unknown variant is present given her disease phenotype. Her parents are currently pregnant and aware of the limitations of testing without knowing this second variant. The literature shows that more contributing mutations for MMA have been identified [Forny et al 2014]; our patient is evidence that there is more to be done.

PrgmNr 3049 - Prevalence and Type of Gastrointestinal Symptoms in the Adult Skeletal Dysplasia Population

View session detail

Author Block: **E. M. Carter**, C. L. Raggio, H. LoTurco, D. McInerney; Hosp. for Special Surgery, New York, NY
Disclosure Block: **E.M. Carter:** None.
Skeletal dysplasias are a heterogeneous group of >450 genetic disorders that affect the size and shape of the skull, trunk, and extremities to varying degrees, and are frequently associated with dwarfism. Clinical observations and patient-reported concerns indicate that gastrointestinal (GI) manifestations are an issue for this population, but quantitative information regarding prevalence and severity of GI symptoms is limited. We examined the frequency and characteristics of GI symptoms in adults with skeletal dysplasias by reviewing their responses to a self-administered questionnaire that measures perceived severity of GI symptoms, the Gastrointestinal Symptom Rating Scale (GSRS). This IRB-approved retrospective review included GSRS responses from 101 adults with a clinical and/or genetic skeletal dysplasia diagnosis. Participant demographics, medication history, and ambulatory status were collected from medical records. Responses from 73 females, 27 males, and 1 transgender man were included. The most common skeletal dysplasias were osteogenesis imperfecta (OI, n=51, 50.5%) and achondroplasia (n=13, 12.9%). Average age was 41.5Â±14 yrs, average height was 136.6Â±22.1 cm, average weight was 58.3Â±20.1 kg, and average BMI was 31.57Â±9.8 kg/m^2. Fifty-five (54.4%) ambulated independently, 23 (22.8%) used an assistive device, and 23 (22.8%) were wheelchair-bound or non-ambulatory. Participants took an average of 3.23Â±2.80 medications, most commonly ibuprofen (n=20, 19.8%) and acetaminophen (n=18, 17.8%). Compared to published GSRS reference data, our cohort scored higher on reflux, diarrhea, and total scores, and lower on abdominal pain and indigestion scores; none reached statistical significance. Though OI respondents had more severe symptoms across all domains, only reflux was significant (p=0.009). Achondroplasia scores were higher for indigestion, constipation, diarrhea, and total scores, and lower on abdominal pain and reflex scores than the general population; only diarrhea was significant (p=0.034). There were no statistically significant differences in any of the domain or total GSRS scores across ambulatory status groups. Height positively correlated with abdominal pain domain score (p=0.033). The number of medications positively correlated with all GSRS scores (p=0.013). Adults with skeletal dysplasias do not report more severe GI symptoms than the general population. Future studies should include larger numbers of individuals with specific skeletal dysplasia diagnoses and investigate the clinical significance between GI symptoms and other quality of life measures in adults with skeletal dysplasias.

PrgmNr 3050 - RNA-seq identifies novel regulatory variants underlying Glycogen Storage Diseases

View session detail

Author Block: **A. Iyengar**[1], X. Zou[2], S. Clarke[1], J. Dai[1], R. Gibson[1], K. Patterson[3], A. Safi[1], D. S. Bali[4], W. Majoros[5], A. S. Allen[6], T. E. Reddy[1], G. E. Crawford[1], P. Kishnani[4], University of Washington Center for Mendelian Genomics; [1]Duke Univ., Durham, NC, [2]durham, NC, [3]Univ of Washington, Seattle, WA, [4]Duke Univ Med Ctr, Durham, NC, [5]Duke Univ., DURHAM, NC, [6]Duke Univ, Durham, NC
Disclosure Block: **A. Iyengar:** None.
Genetic causes for Glycogen Storage Diseases (GSDs) are often unclear after targeted DNA sequencing despite clear clinical and biochemical diagnoses and known causal genes. This has substantial implications for early therapeutic interventions that could improve prognosis. Genetic testing typically assumes that genetic disease is caused by coding mutations, leading to missed non-coding or regulatory pathogenic variants. We aimed to detect such variants in autosomal recessive GSD IX and IIIa patients with a known pathogenic variant on only one allele. Using whole-genome sequencing and targeted PacBio long-read RNA-seq, we identified a novel non-coding variant and mechanism underlying GSD IX (phosphorylase kinase deficiency) in two siblings. Whereas GSD IX patients typically present with A on one allele of *PHKG2*, but the second pathogenic variant was unknown. RNA-seq confirmed the predicted splice event on transcripts from the 1st allele, and also identified inclusion of a 76bp cryptic exon on transcripts from the 2nd allele. Both splice events caused frameshift and a premature stop codon; however, some transcripts from the 2nd allele remained normal, potentially explaining the siblingsâ higher PhK activity. We hypothesized that the 76bp cryptic exon was caused by a leaky cryptic splice site T>G variant (rs1433965292, MAF=0.00002) ~1kb downstream of exon 6 in *PHKG2*. To test this, we used a CRISPR/Cas9 approach to generate T/T, T/G, and G/G HEK293T cell lines. RT-PCR of the mutant (T/G and G/G) lines amplified both the normal transcript and that containing the cryptic exon. G/G lines showed PHKG2 expression, and T/G lines predominantly expressed the normal transcript. This is consistent with nonsense-mediated decay of only the transcripts containing the cryptic exon. A homozygous mutant G/G line showed 25% PhK activity compared to a wild-type T/T line, and experiments are underway to confirm this effect in several independent clones. Additionally, in a patient with GSD IIIa, characterized by debranching enzyme (AGL) deficiency, PacBio RNA-seq identified exon 23 skipping in 32% of *AGL* transcripts. This causes a premature stop codon in exon 24/35. Targeted long-read Oxford Nanopore DNA sequencing identified a 1.5kb deletion including a portion of exon 23, which likely explains exon 23 skipping. Further characterization of these variants and their effects will provide insight into the utility of RNA-seq as a diagnostic tool in the genetics clinic.

PrgmNr 3051 - Spectrum of Germline and Somatic Mitochondrial DNA Variants in Tuberous Sclerosis Complex

View session detail

Author Block: **K. GIANNIKOU**[1], K. Martin[2], A. Abdel-Azim[1], T. R. Hougard[3], Y. Tang[3], J. MacKeigan[2], D. J. Kwiatkowski[1], H. C. Lam[3]; [1]Cancer Genetics Lab., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [2]Coll. of Human Med., Michigan State Univ., East Lansing, MI, [3]Div. of Pulmonary and Critical Care Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA
Disclosure Block: **K. Giannikou:** None.
Background: Tuberous Sclerosis Complex (TSC) is an autosomal dominant genetic disease causing multisystem growth of hamartomatous tumors in brain, heart, skin, lung and kidney and it is known to be due to genetic alterations in either *TSC1 or TSC2*. While the nuclear genome has extensively been studied in TSC, there is limited knowledge regarding the role of mitochondrial DNA (mtDNA) in TSC pathogenesis. **Aim:** To examine the prevalence and spectrum of mtDNA variants in TSC patients and correlate them with clinical features and disease severity since mtDNA variants may act as a disease modifier contributing to tumor development and the remarkable phenotypic heterogeneity seen in TSC. **Methods:** We analyzed mtDNA from buccal swabs from 102 TSC patients (44 male, 54 female, 4 unknown, median age: 31 years; 11 familial cases) using deep coverage amplicon massively parallel sequencing (median read coverage: 7,349). mtDNA analysis was also performed in 100 TSC related tumors (58 kidney angiomyolipoma, 24 SEGA, 11 cortical tubers, 2 LAM, 5 TSC-RCC) with matching normal sample (n=9) from 70 patients; 80 tumors had exome data available. Alterations in mitochondrial copy number were determined by qPCR in tumor and matching normal. **Results:** A median of 21 non-synonymous mtDNA variants were identified in 102 buccal swabs, with high homoplasmy (median: 99.62% allele frequency) mainly missense of unknown significance. A pathogenic variant (UUR;MT-TL1; m.3243A>G, heteroplasmy 12%) was identified in one male TSC patient. Five VUS small indels with >97% heteroplasmy were identified in five individuals. Large mtDNA deletions were not detected. Analysis of TSC tumors demonstrated similar spectrum of mtDNA variants as seen in buccal swabs. mtDNA variants did not correlate with any pathological TSC features. qPCR analysis did not reveal changes in mitochondrial content between tumors and corresponding normal tissue. **Conclusions:** Our study provides insight into the mtDNA landscape of TSC for first time, demonstrating that mtDNA genome is stable within the tumors analyzed and across different tissues.

PrgmNr 3052 - 7 day-old boy presenting with anuria and electrolyte imbalance: Novel, disease-associated variant identified in *SLC5A1* gene

View session detail

Author Block: **E. Beals**[1], P. Bizargity[1,2]; [1]Donald and Barbara Zucker Sch. of Med. at Hofstra/Northwell, Hempstead, NY, [2]Northwell Hlth., New Hyde Park, NY
Disclosure Block: **E. Beals:** None.
The solute-like carrier 5a1 (SLC5A1) gene provides instructions for producing a protein called sodium/glucose cotransporter protein 1 (SGLT1). The SLC5A1 gene is located on chromosome 22q12.3; loss of function mutations in this gene can cause the intestinal monosaccharide transporter deficiency known as glucose and galactose malabsorption syndrome (GGM, OMIM 606824). This rare genetic disorder presents in infancy with abdominal bloating, life-threatening osmotic diarrhea, and dehydration. Here we present a 7-day-old full-term male weighing 7 lb 15 oz, who was transferred from an outside hospital for 2 days of non-bloody diarrhea, hyperkalemia (6.4 mmol/L), and anuria. The patient had previously been breast-feeding, taking formula well, and stooling daily. He was born at term to nonconsanguineous parents and has one healthy older sibling. On initial exam he was lethargic with a soft, nondistended abdomen and no dysmorphic features. His laboratory work up revealed hypoglycemia and a high anion gap metabolic acidosis with pH 7.17, pCO2 31 mmHg, bicarbonate 12 mmol/L, and an anion gap of 18 mmol/L. His lactic acid was 3.1 mmol/L. He was hypernatremic to 183 mmol/L, his blood urea nitrogen was 116 mg/dl, and his creatinine was 4.3 mg/dl. Metabolic work up including plasma amino acids, urine organic acids, and acylcarnitine profile were nondiagnostic. A renal ultrasound showed mildly echogenic kidneys and a voiding cystourethrogram demonstrated normal anatomy. The patient received peritoneal dialysis, electrolyte repletion, airway management with intubation, and total parenteral nutrition. Once he recovered from acute kidney failure, he was extubated and started on oral feeds. The patient had intractable diarrhea with every feeding and his stool was found to be elevated in reducing substances. Endoscopic biopsy of his upper duodenum showed normal histology. Rapid whole exome sequencing revealed a novel, homozygous variant (c.826 T>C, p. W276R) in the SLC5A1 gene, classified as likely pathogenic. Both parents are heterozygous carriers. The patient was started on a carbohydrate-free, fructose-based formula, after which his condition stabilized and he began to gain weight. An estimated 300 individuals are affected by pathogenic mutations in the SLC5A1 gene, with a total of 56 mutations reported. Most patients are the product of consanguinity and present at birth with diarrhea and dehydration. To our knowledge, this variant has not been previously reported as pathogenic nor benign and has not been observed in large population cohorts. In silico analysis supports that this missense variant has deleterious effects on protein structure/function.

PrgmNr 3053 - Ascertainment context is crucial: The penetrance of age-dependent monogenic disease variants depends on ascertainment context

View session detail

Author Block: **U. T. Mirshahi**[1], K. Colclough[2], C. F. Wright[3], A. R. Wood[3], R. Beaumont[3], J. Tyrrell[3], T. Laver[3], R. Stahl[1], A. M. Golden[1], J. Goehringer[1], T. F. Frayling[3], A. T. Hattersley[4], D. J. Carey[5], M. N. Weedon[3], K. A. Patel[3]; [1]Weis Ctr. for Res., Geisinger Clinic, Danville, PA, [2]Royal Devon and Exeter NHS Fndn. Trust, Exeter, United Kingdom, [3]Inst. of BioMed. & Clinical Sci., Univ. of Exeter Med. Sch., Exeter, United Kingdom, [4]Univ. of Exeter Med. Sch., Exeter, United Kingdom, [5]Geisinger Clinic, Danville, PA

Disclosure Block: **U.T. Mirshahi:** None.

BACKGROUND: Accurate penetrance of monogenic disorders is often unknown due to a phenotype-first approach to genetic testing. Here, we use a genotype-first approach in four large cohorts with different ascertainment contexts to accurately estimate penetrance of the three commonest causes of monogenic diabetes, Maturity Onset Diabetes of the Young (MODY). We contrast *HNF1A*-MODY / *HNF4A*-MODY which causes an age-related progressive diabetes and *GCK*-MODY, which causes life-long mild hyperglycemia. **METHODS**: We analyzed clinical and genetic sequencing data from four different cohorts: 1742 probands referred for clinical MODY testing; 2194 family members of the MODY probands; 132,194 individuals from a US healthcare-based cohort; and 198,748 individuals from a UK population-based cohort. **RESULTS**: Age-related penetrance of diabetes for pathogenic variants in *HNF1A* and *HNF4A* was substantially lower in the clinically unselected cohorts compared to clinically referred probands (ranging from 32% to 98% at age 40yrs for *HNF1A,* and 21% to 99% for *HNF4A*). The background rate of diabetes, but not clinical features or variant type, explained the reduced penetrance in the unselected cohorts. In contrast, penetrance of mild hyperglycemia for pathogenic *GCK* variants was similarly high across cohorts (ranging from 89 to 97%) despite substantial variation in the background rates of diabetes. **CONCLUSIONS**: Ascertainment context is crucial when interpreting the consequences of monogenic variants for age-related variably penetrant disorders. This finding has important implications for opportunistic screening during genomic testing.

PrgmNr 3054 - Challenges in Classifying Clinically Relevant Germline *RTEL1* Variants

View session detail

Author Block: **A. S. Thompson**, M. Niewisch, L. J. McReynolds, S. A. Savage; Natl. Cancer Inst., Rockville, MD
Disclosure Block: **A.S. Thompson:** None.
RTEL1 (regulator of telomere elongation helicase 1) is a DNA helicase essential for DNA replication, recombination, and telomere homeostasis. Common intronic *RTEL1* SNPs are associated with brain tumor susceptibility whereas rare coding variants cause telomere biology disorders (TBDs). TBD-associated *RTEL1* variants have been reported in dyskeratosis congenita (DC), Hoyeraal-Hreidarsson syndrome (HH), and familial pulmonary fibrosis and are inherited in autosomal dominant (AD) and autosomal recessive (AR) patterns. Bioinformatic predictions of deleteriousness are often required for *RTEL1* variants because of their rarity and lack of high throughput assays. This results in most variants being classified as variants of uncertain significance (VUS).
Clinical presentations and survival (Kaplan-Meier estimates) were compared between patients with AD or AR *RTEL1* variants (44 affected individuals from 14 families) participating in the National Cancer Instituteâ□□s IRB-approved study of Inherited Bone Marrow Failure Syndromes (NCT00027274). We also performed a comprehensive literature review of disease associated *RTEL1* variants and of rare variants (MAFPatients with AR *RTEL1* variants showed significantly lower overall survival than those with AD *RTEL1* variants (p=0.00048). Of 246 variants in 52 publications, 91 (37%) LP *RTEL1* variants were identified. 42 (17%) disease-associated variants were classified as LB. The remaining disease-associated variants (113, 46%) in the literature were classified as VUS due to MAF >1%, limited clinical data, and/or insufficient bioinformatic predictions. Loss of function (LOF) variants accounted for ~2% of rare *RTEL1* variants present in gnomAD.
Our results suggest that the frequency of rare LOF *RTEL1* variants in the general population may be more common than expected. Patients with AD *RTEL1* genotypes show significantly better survival than AR *RTEL1* genotypes. Functional studies are required to classify the majority of novel disease-associated variants to further understand disease etiology.

PrgmNr 3055 - Colorblindness gene implicated for myopia in highly aggregated Pennsylvania Amish families

View session detail

Author Block: **A. M. Musolf**[1], D. Stambolian[2], J. E. Bailey-Wilson[3]; [1]NIH, Baltimore, MD, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]NIH/NHGRI, Baltimore, MD
Disclosure Block: **A.M. Musolf:** None.
Family studies offer good power to identify highly penetrant genetic variants that are rare in a population but enriched within a family; such studies have experienced a renaissance with the advent of affordable whole genome sequencing (WGS). Myopia (nearsightedness) has become a major health concern, reaching epidemic proportions in some countries. It is also the second leading cause of blindness worldwide. Although myopia is known to be caused by both environmental and genetic factors, its genetic etiology remains unclear. GWAS studies have identified common variants of low to moderate effect associated with myopia, yet much of the heritability is still missing.
This study uses a family-based approach; we performed WGS on 97 individuals from 7 extended Pennsylvania Amish families with prior evidence of linkage to myopia. Founder populations such as the Amish also allow for utilization of exclusive genomic architecture, like unique haplotypes, to better identify potential risk variants. The Amish also have low exposure to some known environmental myopia risk factors.
We performed genetic linkage analysis on these families assuming an autosomal dominant risk allele with 90% penetrance with no phenocopies. We identified 88 genome-wide significant variants across the families, localized to 4q13.1, 5p15.33, 8q21.3, and 9p24.1; 26 of these localized to the *CNGB3* gene at 8q21.3, including the only two exonic variants, which are predicted damaging. *CNGB3* is an excellent candidate as it is expressed in the eye and is causal for both achromatopsia (total colorblindness) and progressive cone dystrophy. Functional analysis of *CNGB3* is currently planned.

PrgmNr 3056 - Comparative analysis of the contribution of copy number and single nucleotide variants to the pathogenesis of idiopathic hypogonadotropic hypogonadism

View session detail

Author Block: **M. Stamou**[1], H. Brand[2], M. Wang[1], I. Wong[2], L. Plummer[1], M. F. Lippincott[1], R. Balasubramanian[1], M. E. Talkowski[2], S. Seminara[1]; [1]Harvard Reproductive Sci. Ctr., The Reproductive Endocrine Unit and The Endocrine Unit of the Dept. of Med., Massachusetts Gen. Hosp., Boston, MA, [2]Ctr. for Genomic Med., Massachusetts Gen. Hosp.; Dept. of Neurology, Massachusetts Gen. Hosp. and Harvard Med. Sch.; Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Boston, MA

Disclosure Block: **M. Stamou:** None.

Introduction: Our group studies Idiopathic Hypogonadotropic Hypogonadism (IHH), a rare form of hypogonadism in which mutations in over 30 genes have been identified as contributing or causing the infertility phenotype. Although identification of single nucleotide variants (SNVs) in these genes has transformed the understanding of the genetics of this condition, the contribution of structural variation to IHH has not been well quantified. In this study, we examined the prevalence of copy number variants (CNVs) in known IHH genes using the worldâ□□s largest cohort of IHH patients. **Methods:** We used exome sequencing (ES) from 1,394 IHH probands to analyze CNVs in 36 known IHH genes using GATK-gCNV, a high sensitivity CNV calling pipeline. We prioritized CNVs with a site frequency Results: A total of 11 IHH genes harbored rare CNVs (size range: 400bp to 9.6 MB) affecting 32/1394 IHH probands (2.3% of the cohort). None of the affected probands carried more than 1 CNV. The CNVs included 24 deletions that disrupted 7 genes (*ANOS1, FGFR1, GNRHR, GNRH1, SEMA3A, TCF12 and NDNF*) and 8 duplications that affected 5 genes (*ANOS1, PROKR2, NSMF, PROP1 and IGSF10*). Half of the deletions detected (12/24) were single gene deletions and the rest were multigenic, including well-known contiguous gene deletion syndromes such as the Xp22.3 contiguous gene deletion syndrome (n= 5). In contrast, most duplications (6/8) encompassed only one single gene. Inheritance was discernible in 10 pedigrees which revealed 4 *de novo* deletions, 5 X-linked recessively inherited deletions and 1 dominantly inherited duplication which did not segregate appropriately. Comparative analysis among the four most prevalent IHH genes (*FGFR1, CHD7, ANOS1* and *GNRHR*) showed that while *ANOS1*, *FGFR1* and *GNRHR* mutational spectrum include both SNVs and CNVs, *CHD7* mutations are exclusively SNVs. **Conclusions:** This comprehensive CNV analysis of known IHH genes in a large IHH cohort reveals only a modest CNV contribution (~2.3%) to IHH pathogenesis. The precise mechanisms through which both deletion and duplication events lead to IHH pathogenesis requires additional investigation. The absence of CNVs in nearly two-thirds of the known IHH genes may relate to strong negative selection at these loci and requires additional follow up evaluation.

PrgmNr 3057 - Density Volumetric Analysis of Cystic Lung Disease in Proteus Syndrome

View session detail

Author Block: **C. Ours**[1], M. Y. Chen[2], J. C. Sapp[1], A. Buser[1], M. Hodges[1], L. G. Biesecker[1]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]Natl. Heart, Lung, and Blood Inst., Bethesda, MD
Disclosure Block: **C. Ours:** None.
Background: Cystic lung disease is the second leading cause of death in Proteus syndrome after venous thromboembolism. This obstructive pulmonary disease is progressive and often complicated by restrictive lung disease secondary to scoliosis and chest wall deformities. Medical management is supportive while surgical approaches such as lobectomy may offer benefit for regionally limited disease. Pulmonary function tests can be helpful to monitor disease progression but have limitations, particularly in younger children. Radiographs and computer-aided diagnostics have been used to quantify other cystic lung diseases such as emphysema and lymphangioleiomyomatosis. We collected computed tomography (CT) images of the lungs to perform density volumetric analysis to quantify cystic lung disease in Proteus syndrome. Methods: Computed tomography images of the lungs were gathered from individuals enrolled on a Natural History Study of Proteus Syndrome (NCT00001403). These included imaging studies performed at the National Institutes of Health Clinical Center and other facilities. Post-operative images were not included. Cystic volumes of the lungs were measured using a computer-aided diagnostic system which algorithmically determines total lung volume and uses an adaptive Hounsfield unit threshold to define cystic volumes. The combined cystic volume was divided by total lung volume to obtain a percentage of cystic volume in the lungs. We defined cystic lung disease as individuals who had greater than 5% cystic volume. A linear mixed-effects model was used to estimate the annual increase in percent of cystic lung in children (0-15 years) and in adolescent and adult individuals (>15 years) with cystic lung disease separately. Results: There were a total of 76 CT imaging studies included from 26 individuals. The median age at earliest imaging was 11.4 years (IQR 10.4-22.7). Serial imaging studies (i.e., two or more) were available from 21 individuals. The median interval between most recent and earliest imaging was 5 years (IQR 2.4-8.4). Cystic lung disease was present in ten individuals (38%). Six individuals with cystic lung disease were less than 15 years old at time of earliest imaging. Four individuals underwent surgical intervention. The annual increase in cystic volume in young children was 6% (95% CI 3.0-9.2%) and in adolescent and adults was 1.1% (95% CI -0.9-2.8%). Conclusion: Cystic lung disease of Proteus syndrome is amenable to quantitation using algorithmic based approaches. Serial measurements demonstrate the progression of cystic volume over time. The rate of progression is greater in children compared to adolescents and adults.

PrgmNr 3058 - Determining the role of Frizzled pathway candidates genes in causing the rare inherited blinding disorder FEVR

View session detail

Author Block: **S. van der Ende**[1], J. M. Robitaille[2], C. R. McMaster[1]; [1]Dalhousie Univ., Halifax, NS, Canada, [2]Dalhousie Univ, Halifax, NS, Canada
Disclosure Block: **S. van der Ende:** None.
Familial exudative vitreoretinopathy (FEVR) is an inherited childhood blinding disorder with an estimated incidence of 1:10,000 affecting approximately 12,500 babies born worldwide per year. The primary cause of FEVR is an inability to vascularize the retina, with vision loss ranging from mild to severe. Secondary complications can occur in a subset of patients; they present unexpectedly and result in irreversible vision loss. Current best practices manage the disease and prevent vision loss, but only if diagnosed before complications arise. A genetic diagnosis is critical for early treatment, however, not all FEVR causative genes are known. Next-generation sequencing was performed on a 126 patient FEVR cohort and variants in known causative genes were identified in 50% of the patients. From the remaining patients, rare *DVL3* variants were identified in 4 separate patients. Effects of the *DVL3* patient-derived variants on known frizzled signalling pathways were investigated using cell based assays. The patient-derived *DVL3* variants were found to affect two different frizzled pathways: the canonical norrin/frizzled4 pathway and the frizzled/Ca2+ pathway. Curiously, the location of the mutations within the DVL3 protein dictates which pathway is affected: the N-terminal half affects norrin/frizzled4 signalling, and the C-terminal half affects the frizzled/Ca2+ signalling pathway. The norrin/frizzled4 pathway is the traditional FEVR mechanism; *DVL3* as a candidate FEVR gene introduces the frizzled/Ca2+ pathway as a new mechanism for FEVR pathogenesis.

PrgmNr 3059 - Diagnostic yield of CNV analysis in primary immunodeficiencies

View session detail

Author Block: **B. Seifert**[1], M. Similuk[1], M. R. Setzer[1], J. Yan[2], M. Kamen[2], C. Jodarski[1], L. Jamal[3], K. Jevtich[2], Y. Yu[1], R. Duncan[1], V. Kuram[1], J. Lack[1], R. Ghosh[1], S. M. Holland[1], L. M. Franco[1], M. A. Walkiewicz[1]; [1]NIH, Bethesda, MD, [2]Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, [3]Natl. Inst. of Allergy and Infectious Diseases/ NIH Clinical Ctr. Dept. of Bioethics, Bethesda, MD
Disclosure Block: **B. Seifert:** None.
Background & Rationale: Primary immunodeficiencies (PID) include disorders affecting different aspects of the immune system. While exome sequencing (ES) can provide molecular diagnoses in many individuals with PID, calling of copy number variants (CNVs) in ES is not reliable due to variation in coverage. Therefore, chromosomal microarray analysis (CMA) can increase the diagnostic yield through detection of clinically relevant copy number variants (CNVs) pertaining to immunological phenotypes. Here, we provide the results of CMA in probands enrolled in the National Institute of Allergy and Infectious Diseases (NIAID) Centralized Sequencing Program at the National Institutes of Health.
Methods: Research-based ES was performed on 1,000 study probands. CNVs were evaluated via CMA on a subset of probands with a syndromic clinical presentation or a single high-priority variant identified by sequencing for a recessive condition. CMA was performed with a custom comparative genome hybridization array designed by Agilent Technologies in collaboration with NIAID, with clinical interpretation completed at Baylor Genetics.
Results: Among 1,000 probands undergoing ES, CMA was performed on 376 probands (37.6%). A clinically relevant CNV was detected in 18 probands (4.8%), and 104 probands (27.7%) harbored a CNV unrelated to the clinical phenotype of the patient. Twenty CNVs were detected in the 18 probands with clinically relevant CNVs. Nine of the twenty CNVs (45.0%) involved genes classified by the International Union of Immunological Societies (IUIS) as associated with inborn errors of immunity, of which 2 (10%) were upgraded in classification due to recently published literature evidence. Probands who received a diagnosis based on CMA results were more often male, younger in age, and more likely to have multiple organ systems affected in comparison to probands in whom a molecular diagnosis could not be established.
Conclusions: CMA had a diagnostic yield of ~5% in individuals with immunological phenotypes, highlighting the contribution of CNV detection to the diagnosis of PID in individuals who may receive an inconclusive or negative result by ES alone. Additionally, over half of the CNVs included genes unrelated to immunological phenotypes, underscoring the complexity of phenotypic presentation in patients with PID.

PrgmNr 3060 - Exome sequencing findings in patients with bronchiectasis within an immune system disorder cohort

View session detail

Author Block: **J. Yan**[1], R. Ghosh[2], M. Similuk[2], V. Kuram[1], J. Lack[1], M. R. Setzer[1], M. Kamen[1], C. Jodarski[3], L. Jamal[4], B. Seifert[5], K. Jevtich[1], Y. Yu[6], R. Duncan[1], D. Hunt[1], A. Borges[1], K. Kumar[1], K. Olivier[1], S. M. Holland[1], M. Walkiewicz[1]; [1]Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, [2]NIH, Bethesda, MD, [34]Natl. Inst. of Allergy and Infectious Diseases/ NIH Clinical Ctr. Dept. of Bioethics, Bethesda, MD, [5]Natl. Inst. of Allergy and Infectious Diseases (NIAID), Bethesda, MD, [6]Ellicott City, MD
Disclosure Block: **J. Yan:** None.
Aims: Bronchiectasis, a permanent dilation of the airways, is a complex heterogeneous disorder. It can occur as a result of opportunistic infections in primary immune deficiencies (PID) and has distinct disease trajectories, ranging from recurrent pulmonary infections in the context of a specific PID to later onset with no known underlying illness. It has been attributed to an interplay between host genetic variation and environmental exposures with studies supporting multigenic etiology. The phenotypic spectrum is wide-ranging in disease subtype and severity, making treatment and understanding disease etiology difficult. Here, we present a framework for studying the molecular basis of bronchiectasis within a mixed cohort of individuals with suspected immune disorders and idiopathic bronchiectasis.
Methods: Individuals enrolled in the National Institute of Allergy and Infectious Diseases Centralized Sequencing Program (CSP) underwent exome sequencing with clinical-grade interpretation. All variants reported to patients were validated by Sanger sequencing.
Results: The CSP exome sequencing cohort included 341 probands whose clinical presentation included bronchiectasis. Of these, 114 (33.4%) were found to have a primary molecular diagnosis consisting of pathogenic or likely pathogenic variants in genes associated with the probandâ□□s immune disorder. Among probands with molecular diagnoses, the most common findings consisted of variants in *CFTR* (n=18, 15.8%), variants in genes related to primary ciliary dyskinesia (n=14, 18.9%), and variants contributing to PIDs, most commonly in *PIK3CD* (n=13, 11.4%), *STAT3* (n=12, 10.5%) and *STAT1* (n=7, 6.1%). The remaining (n=50, 43.9%) molecular diagnoses encompassed a range of immune dysregulation disorders including variants in *CXCR4, FAS, PIK3R1, BTK, CTLA4, AIRE, CARD11, CTNS, DOCK8, FOXN1, HMBS, IL17RC, IL2RG, IRF2BP2, MVK, NKX2-1, RAC2,* and *TNFRSF13B.*
Conclusion: Molecular diagnoses were described in 33.4% of probands with exome sequencing who had bronchiectasis. Findings are consistent with the heterogeneous nature of bronchiectasis and underscore both the multiplicity of PIDs that can result in bronchiectasis, as well as the unknown genetic contributions for the majority cases with bronchiectasis. Burden testing of rare variants contributing to risk of bronchiectasis and assessment of the effect of common variants on disease presentation are warranted to further understand the genetic architecture of bronchiectasis in order to aid in better diagnosis, risk prediction, and treatment.

PrgmNr 3061 - Genomic Analysis of Profound Childhood Hearing Loss in the Yoruba Population of Nigeria

View session detail

Author Block: **R. Faridi**[1], A. Adeyemo[2], P. Chattaraj[1], R. Yousaf[1], R. Tona[1], S. Okorie[2], T. Bharadwaj[3], L. Nouel-Saied[4], A. Acharya[3], I. Schrauwen[5], R. Morell[1], S. M. Leal[6], A. Griffith[7], T. B. Friedman[1], I. Roux[1]; [1]NIDCD/NIH, Bethesda, MD, [2]Inst. of Child Hlth., Coll. of Med., Univ. of Ibadan, Ibadan, Nigeria, [3]Ctr. for Statistical Genetics, Gertrude H. Sergievsky Ctr., Taub Inst. for Alzheimer's Disease and the Aging Brain, Dept. of Neurology, Columbia Univ. Med. Ctr., New York, NY, [4]Ctr. for Statistical Genetics, Gertrude H. Sergievsky Ctr., Taub Inst. for Alzheimer's Disease and the Aging Brain, Dept. of Neurology, Columbia Univ. Med. Ctr.,, New York, NY, [5]Columbia Univ., NYC, NY, [6]Baylor Col Med, Houston, TX, [7]Dept. of Otolaryngology, Coll. of Med., Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN
Disclosure Block: **R. Faridi:** None.
Background There are hundreds of genes in which variant alleles are associated with sensorineural hearing loss in children. However, the causative genes and alleles of hearing loss in Sub-Saharan African populations are still largely unknown. It is widely assumed that there is great heterogeneity of mutations causing hearing loss in Sub-Saharan Africa because of the very old age and ancient admixture of the Sub-Saharan African population. The back-migration of non-Africans into Africa in recent history likely has also added to this heterogeneity. We aimed to further elucidate the variants causing hearing loss in indigenous Sub-Saharan Africans. **Methods** We ascertained 56 small families of Yoruba ethno-lingual ancestry in Ibadan, Nigeria, with at least one individual with non-syndromic, severe to profound, prelingual-onset, bilateral hearing loss that could not be attributed to nongenetic factors associated with hearing loss. We performed a combination of exome and Sanger sequencing analyses of their nuclear and mitochondrial genomes. **Results** No bi-allelic pathogenic variants were identified in *GJB2*, encoding connexin 26, a common cause of deafness in many other populations. Variants predicted to be pathogenic or likely damaging were identified in genes associated with nonsyndromic deafness (*COL11A1, ILDR1*, *MYO15A*, *TMPRSS3,* and *WFS1)*, nonsyndromic deafness or Usher syndrome (*CDH23*, *CIB2*, *MYO7A*, *PCDH15* and *USH2A*), and other syndromic forms of hearing loss (*CHD7, OPA1* and *SPTLC1*). Several rare mitochondrial variants, including m.1555A>G in the gene MT-RNR1, were detected in deaf subjects but not in 118 control Yoruba samples. **Conclusions** Pathogenic variants of *GJB2* play a limited role in the etiology of hearing loss in Sub Saharan African population. Twenty (33%) of 60 independent cases of hearing loss in this cohort of Yoruba families were likely caused by variant alleles of genes reported to underlie deafness in other populations. All identified pathogenic or likely damaging variants in individuals with hearing loss were private, most were detected in compound heterozygosity and 61% had not been previously associated with hearing loss. These results indicate an extremely high level of genetic heterogeneity and underscore the need for more large-scale studies to identify the genetic causes of hearing loss in Africa.

PrgmNr 3062 - Intronic mutations in Puerto Rican (Hispanic) children with rare genetic diseases: Understanding genotype and phenotype correlations at the individual level

View session detail

Author Block: **E. Albino**[1], G. Maldonado-Casalduc[2], W. De Jesus-Rojas[3,4]; [1]Univ. of Puerto Rico - Med. Sci. Campus, Sch. of Hlth.Professions, San Juan, PR, [2]Ponce Hlth.Sci. Univ., Sch. of Med., Dept. of Pediatrics, Ponce, PR, [3]Univ. of Puerto Rico, Med. Sci. Campus, Sch. of Med., Dept. of Pediatrics, San Juan, PR, [4]Ponce Hlth.Sci. Univ., Sch. of Med., Dept. of Pediatrics, San Juan, PR
Disclosure Block: **E. Albino:** None.
Introns are short sequences of genetic material involved in gene regulation, silencing and expression. Hereditary Hemorrhagic Telangiectasis (HHT), Primary Ciliary Dyskinesia (PCD), CFTR-Related Metabolic Syndrome (CRMS) and Ehlers-Danlos (EDS) classical type, are rare genetic diseases with significant associated comorbidities and live-term symptoms and pulmonary complications. Intronic mutations in these genes have been understudied in the Puerto Rican population. We present four pediatric cases with phenotypes associated with these diseases. Case 1 is an 18 y/o male with worsening epistaxis since childhood but normal hemoglobin levels and oxygen saturation. Case 2 is an 8 y/o female with chronic wet cough since birth, bronchiectasis on chest computer tomography and history of recurrent pulmonary infections due to *P. aeruginosa*. Case 3 is a 4 y/o male with persistent bronchial asthma, failure to thrive and intermittent episodes of constipation. Case 4 is a 12 y/o female with marfanoid habitus, joint hypermobility and interstitial lung disease. Cases were referred to a pediatric pulmonologist for evaluation and genetic testing was performed in all individuals. Intronic variations *ENG*, *RSPH4A*, *CFTR* and *COL5A1* were identified and associated with HHT, PCD, CRMS and EDS, respectively. A VUS, c.1428+5C>G (intronic), was identified in *ENG* and it affects a nucleotide within the consensus splice site of the intron. A homozygous, pathogenic variant, c.921+3_921+6del (intronic), was identified in *RSPH4A*. It does not directly change the encoded amino acid sequence but affects several highly conserved nucleotides within the consensus splice site of intron 2. A pathogenic (low penetrance) variant, c.1210-34TG[11]T[5] (intronic), were identified in *CFTR*. A VUS, c.2331+4G>A (intronic), was identified in *COL5A1*. This sequence change falls in intron 26 of the *COL5A1* and it affects a nucleotide within the consensus splice site of the intron. The four pediatric patients present phenotypes associated with the disease. The effect of intronic mutations in the phenotype of rare genetic diseases are not well understood. Access to genetic testing have been able to unmask previously unknown intronic variants in populations with a high degree of both homozygosity and admixture such as Puerto Ricans. Describing these variants and associating them with the phenotypes will help us to understand the pathophysiology of these diseases. Understanding the variants found in our population will contribute to a deeper characterization of the genetic contribution to human phenotypes and disease susceptibility.

PrgmNr 3063 - Mosaicism in a cohort of patients with immune disorders

View session detail

Author Block: **M. Kamen**[1], J. Yan[1], M. Similuk[2], C. Jodarski[3,1], M. Setzer[4], L. Jamal[5], B. Seifert[6], R. Ghosh[2], K. Jevtich[1], Y. Yu[7], R. Duncan[1], S. Holland[1], M. A. Walkiewicz[8]; [1]Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, [2]NIH, Bethesda, MD, [34]Med. Sci. and Computing/NIAID, Bethesda, MD, [5]Natl. Inst. of Allergy and Infectious Diseases/ NIH Clinical Ctr. Dept. of Bioethics, Bethesda, MD, [6]Natl. Inst. of Allergy and Infectious Diseases (NIAID), Bethesda, MD, [7]Ellicott City, MD, [8]Bethesda, MD
Disclosure Block: **M. Kamen:** None.
Background and purpose: Genetic mosaicism describes the phenomenon of an individual having two or more cell populations with different genotypes all originating in a single zygote. While mosaicism is a known feature of cancer and chromosomal aneuploidy, the role of mosaicism in the pathogenicity of immune system disorders is not as well understood.
Objective: This study aims to evaluate our cohort of patients with disorders of immunity to understand the contribution of mosaic variants to disease pathogenesis.
Method: To date, the NIAID Centralized Sequencing Program has performed exome sequencing and issued clinical genetic testing reports on 1,797 families evaluated for immune disorders. All clinically relevant variants were validated by Sanger sequencing. Mosaicism was assessed by alternate allele fraction in exome data.
Results: Within this cohort, we have identified apparently somatic, gonosomal, or germline mosaic variants underlying molecular diagnoses in 30 (1.7%) out of 1,797 families (25 probands and 5 relatives). The variant allele fraction indicative of postzygotic mosaicism ranged from 11%-44%. Possible parental gonosomal mosaicism was detected in three families (0.17%) including unaffected or mildly affected individuals with mosaic variants in *NOD2*, *STAT1*, and *CYBB*. Two families with pathogenic *PIK3CD* variants were highly suggestive of gonadal mosaicism in the parents. Apparent somatic mosaicism in *ASXL1* was identified in three individuals likely due to myelodysplastic syndrome (MDS). One individual with MDS had two mosaic variants of uncertain significance in *ASXL1*. Additionally, our study detected apparently mosaic variants in the following genes related to immune deficiency: *NLRP3* (3), *KIT* (11), *ATM*, *CXCR4*, *IKBKG*, *ITGB2*, *JAK2*, *KRAS*, *IRF8*, *SAMD9L*, *SETBP1*, *TLR8*, and *TRPM4*.
Discussion: The presence of mosaicism in the context of immune deficiencies can have important implications. Genetic counseling of recurrence risk is impacted by the possibility of parental germline or gonosomal mosaicism. Due to technical limitations, the number of cases of mosaicism in this cohort is likely an underestimate. Follow-up studies are warranted to elucidate a more complete understanding of the role of mosaicism in disease severity, which may better inform prognostic and treatment decisions for immune disease.

PrgmNr 3064 - The complexities of germline *PARN* variants in telomere biology disorders and the population

View session detail

Author Block: **M. Thompson**, A. S. Thompson, L. J. McReynolds, S. A. Savage; Natl. Cancer Inst., Bethesda, MD
Disclosure Block: **M. Thompson:** None.
PARN encodes poly(A)-specific ribonuclease, a highly conserved 3â□□ exoribonuclease important in regulating the stability and maturation of RNAs; it acts by shortening the mRNA poly(A) tail through deadenylation. PARN also regulates the turnover of mRNAs and the maturation and stabilization of the RNA component of telomerase (TERC). *PARN* is one of at least 15 different telomere biology genes with clinically significant pathogenic germline variants. Pathogenic germline variants in *PARN* result in reduced levels of TERC by altering its stability and accelerating its degradation. Autosomal dominant and recessive inheritance of rare *PARN* variants have been identified in telomere biology disorders (TBDs) including dyskeratosis congenita (DC), Hoyeraal-Hreidarsson syndrome (HH), and familial pulmonary fibrosis.
To better understand the full scope of germline *PARN* (chr16:14529558-14726585, GRCh37/hg19, ENSG00000140694.12) variation in disease, we conducted a comprehensive literature review, curated *PARN* variants, and assessed evolutionary conservation. The Genome Aggregation Database (gnomAD, N=141,456) was used to identify rare variation (MAF in silico predictors (missense), or there was at least one deleterious prediction by HSF or SpliceAI (splice-site). All frameshift/nonsense variants were classified as deleterious.
There were 362 unique *PARN* variants with MAF PARN variants reported in TBD patients, including 10 splice-site, four intronic, 10 frameshift, 34 missense, seven nonsense, and one synonymous. Our schema classified 40 of the 66 as deleterious and 26 as tolerated. Thirty of the deleterious variants were present in patients with pulmonary fibrosis, five in DC, and five in HH. Thirteen of the 34 missense variants (38%) were at highly conserved residues with the same amino acid present in at least four distinct species. In general, *PARN* is highly conserved with primarily rare variants across all populations evaluated. Notably, commonly used *in silico* predictions of variant deleteriousness were not consistent across literature reported TBD-associated *PARN* variants. Studies to determine the extent to which *PARN* may be constrained are underway.

PrgmNr 3065 - Update to the NIAID Genomic Research Integration System Genomic Analysis Module - Lessons Learned

View session detail

Author Block: **S. Xirasagar**[1], Z. Li[1], J. Barnett[1], S. Ranganathan[1], N. Vlasenko[1], A. J. Oler[1], L. Zhang[1], Y. Fan[1], X. Cheng[1], K. Kaladi[1], E. Karlins[1], J. Lack[2], V. Kuram[2], S. Yazhuk[3], D. Veltri[1], J. Yan[4], M. Walkiewicz[5], R. Ghosh[6], D. Hurt[1]; [1]Bioinformatics and Computational BioSci.s Branch, Office of Cyber Infrastructure and Computational Biology, Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [2]NIAID Collaborative Bioinformatics Resource (NCBR) Frederick Natl. Lab. for Cancer Res., NIH, Frederick, MD, [3]Bioinformatics and Computational BioSci.s Branch, Office of Cyber Infrastructure and Computational Biology, Natl. Inst. of Allergy and Infectious Diseases, NIH, Rockville, MD, [4]Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, [5]Div. of Intramural Res., Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, [6]Div. of Intramural Res., Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, Bethesda, MD

Disclosure Block: **S. Xirasagar:** None.

We report challenges and resolutions in performing a substantive update to the NIAID Genomic Research Integration System (GRIS), a web-based application which integrates patient clinical phenotypes, labs, pedigree information, and exome(WES) and genome(WGS) sequencing data. GRIS is used by molecular geneticists and researchers to explore disease-gene variant relationships. We updated the genetic repository and analysis module, *seqr*, to the latest version to enable variant searching in WGS (as the previous version supported only exomes) data across multiple large datasets. The updated *seqr* also enabled upgrading to the hg38 reference genome, a more accurate version with a richer set of annotations than the hg19 version. We first addressed issues resulting from differences in AWS and GCP Elastic Search (ES) and Spark clusters since *seqr* is customized to use GCP clusters. Since AWS provides different data node specifications and products, we were unable to use the same services. We setup Lambda functions in AWS to build our automated data pipeline. Second, updating to the hg38 version required updating our variant calling, QC, and annotation pipelines. We then mapped variants tagged in the hg19 version to the corresponding variants in hg38 and performed additional manual mapping where variant locations were updated in the newer version. This allowed us to migrate the annotated tags, notes, and audit trail associated with the analysis of these variants. We updated the code to accurately group variants retrieved from multi-project (stored in multiple indexes in the ES cluster) searches across these projects. However, such large-scale searches can return hundreds of thousands of variants, which can be computationally intensive as well as daunting for users. We therefore customized the search to establish thresholds and best practices for searching. We integrated Exomiser which prioritizes likely causative variants based on the variant's predicted pathogenicity, frequency of occurrence in multiple populations, and the degree to which the given phenotype matches the known phenotype of disease genes. GRIS also includes tracks for absence of heterozygosity. In addition to frequencies of variants reported in public genomic databases, frequencies of variants in GRIS patients are displayed. Allele details for the HLA locus genes, which play an important role in the immune system can also be retrieved. Our unique implementation intended for large scale collaborative sequencing and variant research includes over 4000 records linked to WES/WGS and continues to be successfully used by over 200 NIAID Intramural researchers and more recently by other NIH researchers.

PrgmNr 3066 - VEXAS Syndrome Genotype is Associated with Mortality Through Differential Regulation of Non-Canonical Translation of Cytoplasmic UBA1b

View session detail

Author Block: **D. Beck**[1], M. A. Ferrada[1], D. L. Bodian[2], N. T. Strande[3], D. Ospina Cardona[1], L. Wilson[1], P. Hoffmann[1], S. Savic[4], A. K. Ombrello[1], A. Werner[1], D. L. Kastner[1], P. C. Grayson[1]; [1]NIH, Bethesda, MD, [2]Geisinger, North Bethesda, MD, [3]Geisinger, Danville, PA, [4]Univ. of Leeds, Leeds, United Kingdom
Disclosure Block: **D. Beck:** None.
VEXAS or vacuoles, E1 activating enzyme, X-chromosomal, autoinflammatory somatic syndrome is a recently identified genetic syndrome caused by somatic, myeloid restricted, mutations in *UBA1*, which decrease cytoplasmic UBA1 function. Patients with VEXAS syndrome present with a combination of severe inflammatory and hematologic manifestations. We sought to better understand the prevalence, penetrance, and expressivity of pathogenic *UBA1* mutations to better define this disease. Prevalence estimates based on the DiscovEHR cohort predict 1/25,000 individuals carry a pathogenic variant in *UBA1*, with 100% penetrance. Via single center referrals, eighty-eight total patients were identified with pathogenic variants in *UBA1*. Median age at disease onset was 66 years (range 40-85). All patients had hematologic and inflammatory symptoms, were treated with glucocorticoids, and had vacuoles identified on bone marrow biopsy when available. Four disease causing variants were identified including c.118-1 G>C (2%); c.121 A>C, p.Met41Leu (16%); c.121 A>G, p.Met41Val (24%), and c.122 T>C, p.Met41Thr (58%). All mutations were somatic with a median variant allele frequency 72.7% (range 7.5-97%) in peripheral blood. On autopsy material, somatic mutations were restricted to the hematopoietic system and not present in visceral organs. Clinical evaluation of newly diagnosed patients was similar to the discovery cohort with a subset of new manifestations identified including recurrent infections, interstitial lung disease, and ANCA-associated vasculitis. The overall mortality was 27% with a median survival from symptom onset of 10 years. Death was more common in patients with p.Met41Val variant (50%) compared to patients with p.Met41Leu (18%) or p.Met41Thr (22%). UBA1b is the cytoplasmic isoform of UBA1 and is translationally initiated from p.Met41. Translation of the cytoplasmic UBA1b isoform was reduced with all VEXAS mutations, with p.Met41Val expression lowest compared to p.Met41Leu or p.Met41Thr, providing a mechanistic basis for genotype-specific mortality. There was no correlation between mutation burden and survival, but transfusion dependence independently correlated with mortality. Together, our work has defined VEXAS syndrome as a relatively common cause of severe inflammation and bone marrow failure in older adults. We have further uncovered a molecular and genetic correlate for disease severity. Given the high mortality rate and lack of effective medical treatments, patients with VEXAS should be considered for hematopoietic stem cell transplantation, with particular focus on patients with risk factors for increased mortality.

PrgmNr 3067 - 3C minus 2C: X-linked Ritscher-Schinzel/ Cranio-Cerebello-Cardiac syndrome with no cardiac or cerebellar malformation due to missense variant in*CCDC22*

View session detail

Author Block: **S. Basalom**[1], D. Chitayat[2]; [1]Royal Univ. Hosp., Saskatoon, SK, Canada, [2]The Prenatal Diagnosis and Med. Genetics Program, Dept. of Obstetrics and Gynecology and Div. of Clinical and Metabolic Genetics, Dept. of Pediatrics, The Hosp. for Sickkids, Univ. of Toronto, Toronto, Ontario, Canada., Toronto, ON, Canada
Disclosure Block: **S. Basalom:** None.
Ritscher-Schinzel syndrome (RSS) / Cranio - Cerebello- Cardiac syndrome (3C) is a rare syndromic form of intellectual disability associated with posterior fossa defects, cardiac malformations, and minor distal extremities and facial abnormalities. The condition is heterogeneous and recently, two causative genes have been identified - *KIAA0196* and *CCDC22* for Autosomal recessive and X-linked RSS, respectively. Here we report the 8th individual with hemizygous missense variant in *CCDC22*, who presented with severe developmental delay, dysmorphic features, cryptorchidism and oligodontia but with no cardiac or cerebellar involvement.The patient was seen by us at 4 years of age. He was born at term following uncomplicated pregnancy and delivery to a Caucasian mother. He is the eldest in a sibship of two and the parents are healthy and non-consanguineous. At birth he was noted to have bilateral cryptorchidism and nasolacrimal duct obstruction and both were surgically repaired early in infancy. At 4 years of age his weight was 18.7 kg (50th - 85th centile), his length 102.2 cm (15th - 85th centile) and his OFC 52.2 cm (+ 1SD). He had plagiocephaly with a metopic ridge and bitemporal narrowing. He had anterior hair upsweep with a high arched eyebrows, a high forehead with frontal bossing and hypoplasia of the supraorbital ridges. He had malar hypoplasia, a square shaped nose with depressed nasal bridge, narrow nasal root and anteverted nostrils. The eyes were deep-set with bilateral epicanthic folds and puffy eye lids. He had macrostomia, thin upper and lower vermillion, oligodontia and high arched palate. His ears were normal in position but had overfolding of the upper part of the auricles. He had pectus excavatum at the lower part of the sternum. There were webbing between the 2nd-3rd, 3rd -4th and 4th -5th fingers and clinodactyly of the 5th toes on both feet with partial syndactyly of 2nd and 3rd toes with the fourth toe overlapping the 5th toes. He could kick the ball poorly but could jump, skip and hop. He had a short attention span and had poor pincer grip. He had severe language delay, with the expressive language being more severe than the receptive. He was social and affectionate. An echocardiogram showed normal heart and brain MRI at 31 months showed normal cerebellum. WES identified a hemizygous *CCDC22* missense maternally inherited pathogenic variant [c.403C>T (p. Arg135Trp)]. To our best knowledge this is the third family, and eighth reported patient with X-linked recessive intellectual disability with features of RSS/3CC due to missense variant in *CCDC22* without cardiac or cerebellar involvement. Our case suggests changing the clinical spectrum of X-linked RSS/3CC.

PrgmNr 3068 - 45,X/46,XY Mosaicism: Retrospective Study of 77 Patients

View session detail

Author Block: **E. Alkhunaizi**; Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **E. Alkhunaizi:** None.

45,X/46,XY mosaicism is a disorder/ difference of sex development (DSD) posing a great clinical challenge due to the highly heterogeneous spectrum of clinical manifestations, genital abnormalities, gonadal dysgenesis, growth hormone dysregulations, psychological and educational impacts and query increased risk for gonadal malignancies. The full extent of the manifestations in both genders and the long-term outcomes are not fully delineated as previous reports studied small cohorts with limited follow-up.We performed a retrospective chart review of 77 patients with pre- or postnatal diagnosis of 45,X/46,XY mosaicism ascertained from the Hospital of Sick Children and Mount Sinai Hospital in Canada. Short stature was more significant in females than males (p=0.04) and 72% of patients had at least one Turner syndrome stigmata. Growth hormone-treated individuals did not show a significant increase in height compared to the untreated group. All females required puberty induction in contrast to the majority of males. Four females were diagnosed with gonadal tumors while none of the males.

This study represents the largest cohort of patients with 45,X/46,XY mosaicism published so far. It provides a comprehensive overview of the clinical spectrum, long-term outcomes, and the risk of associated tumors. It also highlights our experience with growth hormone therapy and indications for prophylactic gonadectomy.

PrgmNr 3069 - Case report: a boy with imperforate anus and cleft lip

View session detail

Author Block: **M. Park**, M. C. Byler, S. Gupta, S. A. Tatum, A. H. Meier, R. R. Lebel; SUNY Upstate Med. Univ., Syracuse, NY
Disclosure Block: **M. Park:** None.
We report a male child born with imperforate anus and a right incomplete cleft lip with a bifid uvula. The non-consanguineous infant was delivered at 39 weeks of gestation to a 28 year-old G2P1>2 Caucasian female. Prenatal ultrasound had detected the cleft lip. There was no other complication, and no known exposure to teratogens. Neonatal course was notable for discovery of imperforate anus and the patient spent one week in the NICU. Surgical correction of both abnormalities was undertaken. No other dysmorphism was identified. The childâ□□s father has a maternal half-brother who was born with imperforate anus, but there is no family history of facial clefting. Father and paternal grandmother both had reported third-decade hearing deficits, but no dysmorphic features. Chromosome microarray was performed on the patient and both parents. The patient was found to have a 158 kbp duplication at 22q12.3 involving part of the *MYH*9 gene (OMIM #160775). *MYH*9 encodes a heavy chain of non-muscle myosin which participates in cell adhesion, cytoskeleton maintenance, and cytokinesis. Variants of this gene are associated with platelet dysfunction, thrombocytopenia, hearing deficit, cataract and nephropathy, none of which have been observed in our patient. However, it is a candidate gene for non-syndromic cleft lip without cleft palate. *MYH*9-related disorders are caused by loss-of-function variants and inherited in an autosomal dominant (AD) manner. Partial duplication of the gene might have resulted in gene disruption and loss of function. The fatherâ□□s microarray analysis revealed a 161 kbp duplication at 22q12.3 involving all of the *MYH*9 gene, confirming the father as the origin of the duplication recorded in the patient. Additionally, the fatherâ□□s analysis showed a 64 kbp deletion at 6p24.4 involving the *GCNT*2 gene (OMIM #600429). This deletion was not transmitted to the patient, and is considered a variant of uncertain significance (VUS). The motherâ□□s microarray showed a 106 kbp duplication at 3q12.2 involving parts of the *ADGRG*7 (OMIM #612307) and *TFG* (OMIM #602498) genes. *ADGRG*7 encodes G-protein coupled receptor 128 which functions as a transmembrane receptor. *TFG*, also called TRK-fused gene, encodes a protein that participates in endoplasmic reticulum microtubules; it is also involved in the NF-ÎºB signaling pathway. She did not pass this variant, considered a VUS, to the patient.

PrgmNr 3070 - Diagnoses and discoveries from re-analysis of clinical exome sequencing data: The Care4Rare Canada experience

View session detail

Author Block: **T. Hartley**[1], E. Soubry[1], M. Acker[2], B. McInnes[3], A. Salman[4], Care4Rare Canada, D. A. Dyment[5], K. D. Kernohan[6], K. M. Boycott[7]; [1]Children's Hosp. of Eastern Ontario Res. Inst., Ottawa, ON, Canada, [2]The Hosp. for Sick Children, Toronto, ON, Canada, [3]Alberta Children s Hosp./Univ. of Calgary, Calgary, AB, Canada, [4]BC Children's Hosp. Res. Inst., Vancouver, BC, Canada, [5]Ottawa, ON, Canada, [6]CHEO, Ottawa, ON, Canada, [7]Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada
Disclosure Block: **T. Hartley:** None.
OBJECTIVE: Recent studies have shown that the systematic reanalysis of exome sequencing data can yield additional diagnoses in 10-15% of cases. These studies have been limited, however, by small sample sizes (typically less than 50), and they often collect data from a single clinical laboratory. In 2015, Care4Rare Canada offered enrollment to any Canadian patient seen at a participating site with non-diagnostic clinically-performed exome sequencing data. Herein, we sought to report the results for 281 participants whose exome sequencing data had been re-analyzed through our research pipeline.
METHODS: Data was transferred from 11 clinical laboratories and run through a previously described Care4Rare bioinformatics pipeline. For all participants, we collected the original clinical laboratory reports, demographic information, and phenotypic information using PhenoTips. Family-based analysis was performed in collaboration with the referring clinicians and additional research was pursued as appropriate, including â□□matchmakingâ□□, segregation, and splicing studies. All compelling candidates were further examined to determine whether these variants were previously classified by the clinical laboratories, and if not, speculate on why these candidates might have been missed.
RESULTS: The mean age of probands at the time of the reanalysis was 12.5 years old (range 0 to 74) and they presented with a broad range of phenotypes; approximately half of our cohort had syndromic ID. The mean time between clinical analysis and our re-analysis was 20.9 months. Re-analysis yielded 15 diagnoses in known disease genes that fit the patientâ□□s presentation (5%). Of these, 4 of the diagnoses could have been made at the time of initial analysis and had been missed by the clinical laboratoryâ□□s workflow. Another 5 were made based on new genomic knowledge (newly curated disease gene-associations or published mutations). An additional 32 (11%) probands had compelling VUSs in known disease genes that the Medical Geneticist considered to be likely the cause of their patientâ□□s disease but required additional evidence to meet ACMG criteria for pathogenicity. Forty five (22%) had compelling variants in genes that are not yet associated with human disease. Of these, 28 (10% of total cases) have been resolved with patient matchmaking and represent novel disease-gene associations.
CONCLUSIONS: This study suggests there is significant utility in the re-analysis of non-diagnostic exome sequencing data for the identification of both molecular diagnoses (5-16%) and discoveries (10-22%).

PrgmNr 3071 - Exome CNV calling and analysis in a large cohort of families with undiagnosed rare genetic disease

View session detail

Author Block: **G. Lemire**[1,2], E. Evangelista[1], I. Wong[1], E. England[1,2], K. Chao[1], L. Pais[1,2], J. Goodrich[1], C. Austin-Tse[1], S. DiTroia[1], E. O'Heir[1,2], J. Fu[1], Broad CMG Collaborators, M. Babadi[1], S. Lee[1], A. N. Smirnov[1], L. Gauthier[1], S. Baxter[1], D. MacArthur[3,1], H. Rehm[1], M. Talkowski[1], H. Brand[1], A. O'Donnell-Luria[1,2]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Boston Childrenâ□□s Hosp., Boston, MA, [3]Murdoch Childrenâ□□s Res. Inst., Melbourne, Australia

Disclosure Block: **G. Lemire:** None.

The Center for Mendelian Genomics (CMG) at the Broad Institute has sequenced 7,719 families with a suspected genetic disease since 2016. Many had a chromosomal microarray and gene panel sequencing for known causes of disease prior to exome sequencing through the CMG. For typical rare variant analysis, exome sequencing data can be used to call SNVs and indels smaller than 50 base pairs, and standard chromosomal microarrays will detect CNVs larger than 50 kilobases. However, mid-sized CNVs are not detected in routine exome analysis of SNVs and indels but molecular diagnostic laboratories are increasingly including CNV calling on exome data in analysis. Detecting CNVs from exome data has been notoriously difficult due in part to the non-uniform distribution of captured reads secondary to biases introduced by PCR and capture steps and many different algorithms have been developed for this purpose. The Genome Analysis Toolkitâ□□s (GATK) CNV tool, the Germline CNV (gCNV) caller, uses a probabilistic framework to infer rare CNVs from read depth data in the presence of systematic bias. We used the gCNV algorithm to call CNVs across the Broad CMG cohort. While analysis is ongoing, we have diagnosed 138 previously unsolved families to date. The identified CNVs consisted of 109 deletions, 20 duplications and 9 complex CNVs. A CNV in a known gene that is consistent with the phenotype was identified in 114 families and a CNV in a novel candidate gene was identified in 24 families. Supporting genetic and/or experimental evidence were required to consider a CNV in a novel gene as the diagnosis in a given family, most often by additional families identified through Matchmaker Exchange. The predominant phenotype present in these families were neurodevelopmental disorders (67%) followed by neuromuscular disorders (19%). We estimate that about 50% of the CNVs that solved CMG cases would not have been detected by standard chromosomal microarrays. Calling CNVs from existing exome data increases the diagnostic yield for individuals that remain undiagnosed after standard testing approaches, providing a higher resolution alternative to arrays at a fraction of the cost of genome sequencing.

PrgmNr 3072 - Phenotypic spectrum of Down syndrome in 4,209 patients by profiling electronic health records

View session detail

Author Block: **J. Havrilla**[1], M. Zhao[2], I. Helbig[1], E. J. Bhoj[1,3], K. Wang[1,3]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Childrens Hosp. of Philadelphia, Philadelphia, PA, [3]Univ. of Pennsylvania, Philadelphia, PA

Disclosure Block: **J. Havrilla:** None.

Down syndrome (DS) is by far the most common and best known chromosomal disorder in humans, occurring one in every 700 newborns. It is a highly heterogeneous disease, with a wide range of phenotypic variation among affected individuals and diverse comorbidities. We sought to determine if we can characterize the phenotypic spectrum and subgroups of this highly heterogeneous disorder, through natural language processing (NLP) of electronic health records (EHRs).We examined physiciansâ clinical notes from the EHRs of 4,209 DS patients at the Childrenâs Hospital of Philadelphia and 7,845 controls. We mapped clinical data to Human Phenotype Ontology (HPO) terms using MetaMap and catalogued the most critical HPO terms related to DS patients, stratified by odds ratio as compared to controls, with appropriate adjustment of information content. We further examined associated HPO terms by age, in an attempt to surmise the typical age of diagnosis for prevailing traits, as well as minimum ages for diagnosis for traits such as speech delay, which cannot be diagnosed at the beginning of life. We also inferred higher-level, less specific HPO terms from our data to derive larger phenotypic groups.Using these tools, we confirm known clinical features in DS, including congenital heart disease, where we show a ventricular septal defect in 41% of all cases, which we find to be a leading cause of death during the first years of life. We demonstrate cortical cataracts in 42% of DS patients, heterotropia and strabismus in 40% of DS patients, and hearing impairment is prevalent in 51% of DS patients. These results suggest that our NLP-based analysis of EHRs can reproduce well known knowledge. Additionally, we also find lesser known characteristics of DS patients as we attempt to create a phenotypic spectrum to characterize the disorder and its subtypes. For example, we identified several rare, clinical features that are over-represented in DS but are not previously described, such as oliguria and miosis, both with odds ratio over 100.In summary, our analysis profiled the spectrum of phenotypic features in patients with DS in the form of standardized terminologies; we demonstrated that an NLP-based approach has face value and provided a quantitative measure of the HPO terms to facilitate future construction of phenotype-based patient sub-classification models, which will allow for clinical decision support and learning health systems.

PrgmNr 3073 - A partial Xq duplication in a newborn girl: A Case Report

View session detail

Author Block: **S. Mikkilineni**[1], M. Rahman[2], F. Shakil[1], C. KUMAR[1]; [1]Westchester Med. Ctr., Valhalla, NY, [2]WMC, Valhalla, NY
Disclosure Block: **S. Mikkilineni:** None.
Partial duplication of the long (q) arm of chromosome X is an infrequent occurrence with variable presentation. In essence, females harboring the dup(Xq) are less affected than males harboring the same duplication since the abnormal X chromosome is often preferentially inactivated in females. Females with dup(Xq) may exhibit short stature, variable dysmorphic features, developmental delay, hypotonia, feeding problems, hypomyelination, gonadal dysgenesis, secondary amennorhea, intellectual disabilities, poor language and mild scoliosis. Here we present a new case of a pre-term female baby born with intra-uterine growth retardation at 36 weeks 6 days by C-section. The infant was small for gestational age (SGA) born with a birth weight of 1.53 kg (

PrgmNr 3074 - A rare case of multiple mosaic marker chromosomes in a neonate with congenital diaphragmatic hernia and ventricular septal defect

View session detail

Author Block: **L. Schultz-Rogers**, K. Kaiser-Rogers, T. Felton, M. Gucsavas-Calikoglu, L. Ramkissoon; UNC Chapel Hill, Chapel Hill, NC
Disclosure Block: **L. Schultz-Rogers:** None.
Marker chromosomes are abnormal chromosomes that lack sufficient banding to be unambiguously identified by traditional karyotype analysis. Here we present a female neonate referred for karyotype and chromosome microarray analysis (CMA) due to congenital diaphragmatic hernia (CDH) and ventricular septal defect (VSD). G-band analysis of newborn cord blood demonstrated four cell lines containing 1-4 marker ring chromosomes (47~50,XX,+1~4r) and the absence of any 46,XX cells. CMA determined that the marker chromosomes were derived from chr. 6,11,13 and 20. The patterns from chr. 6, 11 and 13 were consistent with a ring as duplicated pericentric material from both the short and long arms was present. The pattern of triplicated material from chr. 20 suggests a more complex rearrangement as it contains two discontinuous interstitial regions from the short and long arm. FISH testing with centromeric chr. 6 and 11 probes or locus-specific probes targeting chr. 13 and chr. 20p11.21 and 20q13.2 confirmed the presence of four different markers and demonstrated they were distributed in various combinations. Together with the CMA and FISH, this patientâ□□s karyotype was described as 47~50,XX,+1~4r.ish r(6)(D6Z1+),r(11)(D11Z1+),r(13)(RP11-69J15+),r(20)(RP11-946M12 enh, RP11-381C11 enh). nuc ish(D6Z1,D11Z1,RP11-69J15)x2~3,(RP11-946M12, RP11-381C11)x2~4[200]. As parental testing was not performed, the inheritance of the markers is unknown, however 70% of marker chromosomes occur *de novo* (Liehr 2021). While the duplicated regions do not contain known triplosensitive regions per ClinGen, nor are there similarly reported duplications noted in the DECIPHER database, it is likely the multiple mosaic marker chromosomes identified in this patient are related to their phenotype. There have been 82 reported individuals with multiple marker chromosomes, and only 21 reported 4 or more markers making the findings in this patient exceedingly rare (Liehr 2021). Complex marker chromosomes such as the chr. 20 marker identified in this patient are also unusual and are predicted to constitute ~8% of identified markers (Liehr 2013). Individuals with â□¥4 markers have been reported with a wide range of phenotypes including developmental delay, hypotonia, dysmorphisms, and heart defects (Liehr 2021). The clinical presentation associated with the presence of mosaicism for multiple small marker chromosomes remains difficult to predict and will depend on the distribution of the cell lines within the tissues of the affected individual. This case highlights the importance of obtaining cytogenetic testing on patients with non-specific congenital abnormalities such as CDH and VSD.

PrgmNr 3075 - Assessment of Alzheimer Disease Related Plasma Biomarker Phosphorylated Tau 181 in Individuals of Diverse Ancestral Backgrounds

View session detail

Author Block: **A. J. Griswold**[1,2], F. Rajabli[1], C. Garcia-Serje[1], J. Arvizu[1], E. Larkin-Gero[1], P. L. Whitehead[1], K. L. Hamilton-Nelson[1], L. Adams[1], M. Contreras[1], S. Tejada[1], P. R. Mena[1], T. D. Starks[3], C. Silva[4], M. L. Cuccaro[1,2], J. M. Vance[1,2], B. Feliciano-Astacio[4], G. S. Byrd[3], G. W. Beecham[1,2], J. L. Haines[5], M. A. Pericak-Vance[1,2]; [1]John P. Hussman Inst. for Human Genomics, Univ. of Miami, Miami, FL, [2]Dr. John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami, Miami, FL, [3]Maya Angelou Ctr. for Hlth.Equity, Wake Forest Univ., Winston-Salem, NC, [4]Univ. Central del Caribe, Bayamon, PR, [5]Case Western Reserve Univ, Cleveland, OH

Disclosure Block: **A.J. Griswold:** None.

Background. The use of plasma proteins as biomarkers for the differential diagnosis of Alzheimer disease (AD), and identification of preclinical AD, is well supported. Particularly, phosphorylated threonine 181 of the Tau protein (pTau181) is significantly higher in individuals with both clinical and neuropathologically confirmed AD relative to cognitively and pathologically intact age-matched controls. However, these observations have been made nearly exclusively in individuals of non-Hispanic European ancestry. Given notable differences in the genetic risk for AD across diverse African and Amerindian ancestries, and the lack of biomarker studies in these populations, generalizability of these findings is not assured. Therefore, the aim of this study is to explore the utility of pTau181 in discriminating clinically diagnosed AD from cognitively intact controls in ancestrally diverse African American and Puerto Rican cohorts. **Methods.** We measured pTau181 concentration in a cohort of 536 African American individuals (140 AD, 396 cognitively intact age-matched controls), 357 Puerto Rican individuals (165 AD, 192 controls), and 56 autopsy confirmed AD cases of European ancestry. Plasma was isolated from EDTA blood tubes using centrifugation and pTau181 measured using the pTau181 Advantage V2 Simoa chemistry assay on the Quanterix HD-X. Samples were randomized in plates, measurements performed in duplicate, and non-parametric Wilcoxon rank sum tests with Bonferroni multiple-comparisons p-value correction were used to test differences in pTau181 concentration. **Results.** Overall, the mean plasma pTau levels in AD cases was higher than controls in African Americans (2.51Â±1.47pg/mL vs 1.23Â±0.73pg/mL, $p_{corr}=7.9x10^{-25}$) and Puerto Ricans (2.39Â±1.19pg/mL vs 1.48Â±1.03pg/mL, $p_{corr}=4.5x10^{-16}$). Importantly, the pTau levels in the European ancestry autopsy confirmed cases (3.08Â±1.37pg/mL) were not significantly higher than in African American ($p_{corr}=0.10$) or Puerto Rican ($p_{corr}=0.26$) cases. **Discussion.** This study suggests that previous pTau181 findings will be generalizable and useful for all individuals, regardless of ancestry. Ongoing measurement and analysis of pTau181 and other AD biomarkers in these and other ancestrally diverse cohorts is currently underway to further expand these findings. Combining genomic and biomarker analyses in diverse individuals will ultimately help understand the underlying genetic risk and to refine clinical diagnoses in individuals of diverse ancestral backgrounds.

PrgmNr 3076 - Clinical utility of long read sequencing in resolving inconclusive diagnoses due to gene conversions and unknown phase

View session detail

Author Block: **R. Rajagopalan**[1], D. A. McEldrew[1], M. Luo[2], L. K. Conlin[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]The Children S Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **R. Rajagopalan:** None.
Short read next-generation sequencing (NGS) has revolutionized genomic diagnostics, but has inherent limitations in detecting variants in difficult to sequence regions (e.g. low complexity, repeats, and segmental duplications).There are several known disease-causing genes that are challenging to sequence using NGS, and thus, the spectrum of benign and pathogenic variation in these genes are not entirely known. Additionally, for autosomal recessive diseases, phasing of two variants generally cannot be determined without parental testing, which are often not available, leading to inconclusive test results. PacBio HiFi and Oxford Nanopore sequencing can produce very long reads (20kb to several hundred kb) with high accuracy, which will resolve many of these challenges faced by NGS. We chose three patients with nonsyndromic hearing loss and potentially diagnostic but inconclusive variants identified by an NGS panel. Causality of these variants could not be confirmed due to the unknown chromosomal phase, and in two cases additionally complicated by the presence of a pseudogene. We performed whole genome PacBio long read DNA sequencing to help in the interpretation of these variants. Patient 1 had a pathogenic two-exon deletion and a variant of uncertain significance in the gene *TRIOBP*. Patient 2 had multiple variants across exon 25 and intron 26, suggesting a 21kb gene conversion event between *STRC* and its pseudogene *STRCP1* along with a single nucleotide variant of uncertain significance in exon 8. Patient 3 had multiple variants in exons 23-25 on both haplotypes suggesting multiple gene conversions between *STRC* and *STRCP1*. Determining phase information from short read exome sequencing data is impossible as the variants in *TRIOBP* were 13kb away from each other, and the variants in *STRC* in patient 2 were 19kb apart. PacBio sequencing and long-read based phasing revealed that the two variants in the gene *TRIOBP* for Patient 1 occurred on the same allele and thus were not diagnostic. For both Patients 2 and 3, PacBio revealed that variants in *STRC* occurred on different alleles and confirmed size and extent of each gene conversion event and confirmed biallelic pathogenic variants for both patients. PacBio long-reads enabled the identification of the gene conversion event without having to perform additional assays and the variant backed phasing resolved the haplotypes readily. Results from these three patients support the significant promise of long-read sequencing in resolving some of the very challenging and frustrating issues in identifying and phasing complex variants.

PrgmNr 3077 - Comprehensive genetic testing of sensorineural hearing loss in children: understanding phenotypic factors that influence the diagnostic yield

View session detail

Author Block: **N. Yamamoto**[1,2,3], J. Balciuniene[1], T. Hartman[1,2], B. Devkota[1,2], E. Bedoukian[1,2], N. Golenberg[1], M. DÃaz-Miranda[1], S. Charles[1], M. C. Dulik[1], J. Germiller[1,4], L. Conlin[1], M. Luo[1,4], I. Krantz[1,2,4]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Roberts Individualized Med. Genetics Ctr. (RIMGC), Children's Hosp. of Philadelphia, Philadelphia, PA, [3]Natl. Inst. of Sensory Organs, Natl. Hosp. Organization Tokyo Med. Ctr., Tokyo, Japan, [4]Perelman Sch. of Med. at the Univ. of Pennsylvania, Philadelphia, PA

Disclosure Block: **N. Yamamoto:** None.

Hearing loss (HL) is a common and heterogeneous sensory disorder in children with a genetic etiology suspected in >60%. Genetic testing is recommended for children with bilateral sensorineural HL (SNHL), which may lead to early interventions. Comprehensive genetic testing (CGT) panels allow interrogation of preselected genes. To evaluate the factors impacting diagnostic yield for children with SNHL, we analyzed the relationship between the genetic findings and clinical information.This study analyzed 474 unrelated probands tested with a tiered exome sequencing (ES)-based CGT at our institute. The CGT interrogated 121 SNHL-related genes for sequence and copy number variations (CNVs) if the first tier (*GJB2* testing) was negative. A total of 1147 reportable variants involving 119 genes were identified in 439 of 474 probands. Overall diagnostic yield was 44% (207/474), and 4 probands carried two genetic etiologies. CNVs or gene conversions accounted for 14% (45/327) of all causative variants, 71% of which were detected in *STRC*. Mutations in *GJB2* were the most common findings (37%, 78/211) leading to a wide range of phenotypes from congenital profound HL to postlingual mild HL, but were much less prevalent in some ethnic groups (13% in African Americans and 0% in Hispanics). The remaining 63% (133/211) involved 40 genes (most commonly *STRC* followed by *MYO15A)*. Phenotypic stratification showed that the diagnostic yield was higher in probands with congenital, severe or profound, flat, and symmetric HL, and lower in probands with postlingual, mild, ascending, unilateral HL, cochlear nerve deficiency, preterm birth, NICU admittance, and developmental delay. Among undiagnosed probands, 19% (50/267) had inconclusive findings due to partial phenotype overlap and/or limited variant evidence, and 56% (149/267) had a single heterozygous variant in a gene associated with autosomal recessive HL, of which 22% (33/149) had phenotypes consistent with the condition related to that genetic etiology. Reflex to ES was performed on 46% (122/267) of undiagnosed probands, and 4% (5/122) had diagnostic findings in 4 genes that were not included on our initial CGT panel (3 syndromic genes and 1 nonsyndromic gene). Two of these 4 genes were included in updated versions of the panel.Our relatively high diagnostic yield with heterogeneous results suggests that a tiered CGT panel provides an efficient way to diagnosis genetic causes of diverse childhood-onset SNHL. While the additional yield of ES was low it may be indicated in some undiagnosed probands in order to capture the broad spectrum of syndromic genes associated with SNHL as well as novel genes not included in CGT panels.

PrgmNr 3078 - Genome-wide Sequencing Ontario (GSO): An implementation pilot to improve rare disease diagnostics

View session detail

Author Block: **M. Gillespie**[1], R. Z. Hayeems[2], C. R. Marshall[2], A. Szuto[2], C. Chisholm[3], D. J. Stavropoulos[2], E. M. Price[3], R. Khan[2], W. Lee[2], L. Huang[3], O. Jarinova[3], R. Mendoza-Londono[2], S. L. Sawyer[4], M. J. Somerville[2], K. M. Boycott[1]; [1]Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, [2]The Hosp. for Sick Children, Toronto, ON, Canada, [3]CHEO, Ottawa, ON, Canada, [4]Children's Hosp. of Eastern Ontario Res. Inst., Ottawa, ON, Canada
Disclosure Block: **M. Gillespie:** None.
Purpose: The Genome-wide Sequencing Ontario (GSO) pilot implementation project has developed an innovative, harmonized, and multi-institutional model for delivering and monitoring the performance of clinical genome-wide sequencing (GWS) in the forms of exome and genome sequencing. The first of its kind in Canada, the GSO pilot will collect data on performance, workflow and implementation outcomes to inform repatriation decisions in Ontario and serve as a model for jurisdictions for whom clinical GWS is a policy priority.
Methods: The GSO implementation pilot is using a hub-and-spoke model to provide clinical GWS to Ontarians. Sequencing and bioinformatics will be performed at SickKids, with analysis and reporting at both SickKids and CHEO (2/3 and1/3 of provincial volume, respectively). Guided by the principles of hybrid implementation-effectiveness study design, two prospective cohorts of patients will be enrolled over a two-year period. Cases enrolled in the implementation cohort (n=1320 singletons, duos, trios and quads drawn from peripheral hospital sites in Ontario) are offered exome sequencing (ES) to optimize the delivery of the current standard of care in Ontario. Cases enrolled in the evaluation cohort (n=650 trios from CHEO and SickKids) are randomized to receive exome or genome sequencing (GS). Collected data will include: patient age, phenotype, ethnicity, consanguinity, molecular testing history, sample processing and analysis time, number and characteristics of primary and secondary variants. Costs for each input related to the laboratory analysis will be calculated by multiplying resource use by unit price. For each additional patient with a pathogenic variant detected, we will determine the incremental costs of GS compared to ES. Descriptive statistics will be used to analyze process and outcome data. Point estimates for diagnostic utility and timeliness will be compared statistically for ES and GS.
Results: Since the launch of the GSO implementation pilot on April 1st 2021, 95 individuals or families with suspected rare genetic diseases have been submitted to the SickKids and CHEO laboratories to undergo clinical GWS. Sequencing has been initiated for 67 families and completed in 46 families.
Conclusion: GSO has developed a centralized approach for sequencing and bioinformatics coupled with a distributed, institution-based mechanism for data analysis, variant interpretation, and reporting. Process and outcome data gathered over the course of the two-year pilot will inform provincial and cross-provincial policy related to the long-term organization, delivery, and reimbursement of genome diagnostics.

PrgmNr 3079 - Going above and beyond in discovering *STRC* pathogenic variants in non-syndromic hearing loss: itâ□□s just the tip of the iceberg

View session detail

Author Block: **J. Balciuniene**[1], M. Diaz-Miranda[1], E. Bedoukian[2], T. R. Hartman[2], I. D. Krantz[3], L. K. Conlin[3], M. Luo[3]; [1]Div. of Genomic Diagnostics, The Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Div. of Human Genetics, Dept. of Pediatrics, The Children's Hosp. of Philadelphia, Philadelphia, PA, [3]The Children's Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **J. Balciuniene:** None.
Pathogenic *STRC* variants cause DFNB16, a common type of childhood-onset hearing loss (HL). *STRC* is located in a segmental duplication region which includes a highly homologous pseudogene, *STRCP1*, and is prone to non-allelic homologous recombination leading to copy number (CN) changes, gene conversions, and fusion genes. Large biallelic deletions involving *STRC* are the most commonly reported pathogenic findings, while sequence variants (e.g. SNVs) are reported in a minority of cases due to genetic testing challenges caused by the presence of *STRCP1*. This has limited our understanding of the full extent of the diagnostic contribution of *STRC* to HL.
The Audiome is a comprehensive panel for genetic testing of pediatric non-syndromic HL, which uses exome-based NGS and array CGH platforms. In addition, it includes a *STRC*-specific module involving long range PCR followed by NGS to identify SNVs, and droplet digital PCR to identify CN changes. Final interpretations are based on integrative analysis of both *STRC*-specific approaches.
Audiome testing of 479 unrelated probands yielded diagnostic findings in 45% (216/479). *STRC* positive findings were identified in 5.6% (27/479) accounting for 12.5% of all diagnoses. Only 33% (9/27) of the *STRC* diagnoses were due to biallelic *STRC* deletions. A heterozygous *STRC* deletion and a hemizygous SNV accounted for 26% (7/27) and compound heterozygous SNVs were identified in 22% (6/27) of *STRC* diagnoses. In two of these patients, several rare heterozygous SNVs originating from *STRCP1* were found in multiple exons, suggesting a large gene conversion event. More complex *STRC* findings were identified in the remaining 18.5% (5/27) *STRC* patients. Two individuals carried a deletion on one allele and a non-functional *STRC-STRCP1* fusion on another allele. In three individuals, an apparently homozygous pathogenic variant was identified by NGS; however, the integrated analysis was consistent with the presence of *STRC* deletion on one allele, and *STRC* duplication on the other allele with one *STRC* copy carrying a pathogenic SNV, and the second copy being a non-functional *STRC-STRCP1* fusion.
While some pathogenic variants involving *STRC* can be detected by standard NGS and array CGH, the inclusion of this *STRC*-specific testing module resulted in a doubling of the number of *STRC*-related diagnoses. Without the *STRC*-specific testing, diagnoses would have been completely missed in 5 individuals, and 8 individuals would have yielded inconclusive results. Our integrated analysis helped to uncover the signatures of complex genomic changes and their potential mechanisms and could serve as a model for analyzing other similar gene regions.

PrgmNr 3080 - Molecular diagnosis as an incidental finding from preconception carrier screening

View session detail

Author Block: **J. Reiner**, L. Rosenblum, W. Xin, Z. Zhou, H. Zhu, N. Leach; Labcorp Genetics, Lab. Corp. of America, Westborough, MA
Disclosure Block: **J. Reiner:** Salary/Employment; Laboratory Corporation of America.
The discovery of incidental or secondary findings is not unique to genetic testing; however, the application of next-generation sequencing (NGS) methods to broad population screening programs presents new opportunities to detect clinically-significant findings unrelated to the intended scope of testing. In the context of preconception carrier screening, the detection of pathogenic variants in either a homozygous or compound heterozygous state may represent an incidental disclosure of a disease status unknown to the patient or referring physician, particularly for conditions associated with reduced penetrance or variable presentation. To investigate the frequency of cases reported with a molecular diagnosis, 73,885 sequencing results from three carrier screening panels were reviewed to identify cases harboring at least two pathogenic or likely pathogenic variants in the same gene with demonstrated phasing in the trans configuration. A total of 464 such cases were identified. The variants observed within this cohort involved 13 genes and are predicted to result in at least 16 distinct Mendelian disorders of varying severity. The majority of these cases (398/464, 86%) had variants associated with asymptomatic or milder disease presentations. Excluding alpha thalassemia trait (n=300) and Duarte galactosemia homozygosity (n=93) resulted in a subset of cases (66/71; 93%) enriched for deleterious variants associated with moderate to profound, highly-penetrant conditions. These data indicate that a small percentage of individuals undergoing preconception screening will have an incidental finding that may require clinical evaluation.

PrgmNr 3081 - Non-invasive diagnostic testing for uterine fibroids: A cell-free DNA approach

View session detail

Author Block: **C. Lee**[1], G. Rivera-Cruz[2], L. J. Johnson[1], M. Seshan[1], R. M. Anchan[1], D. R. Armant[3], C. C. Morton[4]; [1]Brigham and Women's Hosp., Boston, MA, [2]Boston Children's Hosp., Boston, MA, [3]Cradle Genomics, San Diego, CA, [4]Brigham & Women's Hosp, Boston, MA
Disclosure Block: **C. Lee:** None.
Leiomyomas, or uterine fibroids (UF), are diagnosed in almost 70% of self-reported white women and over 80% of self-reported black women by the age 50. Approximately 25% of UF are symptomatic, causing heavy menstrual bleeding and anemia, pain, infertility and recurrent pregnancy loss. Therefore, UF are the leading cause of morbidity and hospitalization for benign gynecological issues among reproductive-aged women. The recent bill introduced in the Senate last summer by Senator Kamala Harris, S.4397: Uterine Fibroid Research and Education Act of 2020, has empowered research institutes to advance fibroid research.
Uterine fibroids are often diagnosed by ultrasound, the current non-invasive diagnostic tool that has its own limitations, especially when deciding the appropriate management. The possibility of uterine leiomyosarcoma (ULMS) among UF further complicates disease management and emphasizes the need for a high quality non-invasive diagnostic tool.
Advances in sequencing technologies and understanding of the genetic/genomic landscape of UF are providing new tools to address this knowledge gap and develop robust diagnostics for UF. Heterozygous somatic mutations in *MED12* are present in 70% of UF. Utilizing cell-free DNA (cfDNA) approaches as a non-invasive diagnostic tool has steadily been gaining attraction especially in the prenatal and oncology setting. Liquid biopsies containing cfDNA shed from cancer cells can harbor genetic biomarkers that aid in not only diagnosing a patientâ□□s neoplasm but also in disease management and prognosis. *MED12* somatic variants associated with UF can be assessed in rare cells present in these accessible specimens by deep sequencing technologies. Compared to peripheral venous blood, cells shed from UF are more concentrated in effluents from the uterus, including menstrual blood (MB) and cervical fluid (CF) and may be identified utilizing *MED12* variants as a valuable biomarker.
In this study, we are recruiting women of reproductive age diagnosed with uterine fibroids who are scheduled for surgery (n=100). IRB approval has been obtained. We will collect cfDNA from various biological samples utilizing a menstrual disc with the aim to detect common *MED12* variants. This novel idea highlights the non-invasive approach of diagnosing uterine fibroids. We will compare the data with women without a medical history of fibroids, endometriosis, or adenomyosis (n=30). A cfDNA approach could overcome limitations posed by current diagnostic tools and contribute to innovative precision medicine approaches with a quality of life impact for management of women with UF.

PrgmNr 3082 - Optical Genome Mapping for Constitutional Postnatal SV, CNV, and Repeat Array Sizing: A Multisite Clinical Validation Study

View session detail

Author Block: **N. Sahajpal**[1], V. Rodriguez[2], L. Kanyo[2], A. A. Stence[3], J. R. Hackman[3], A. Mondal[1], A. Bossler[3], P. L. Nagy[2], R. Kolhe[1]; [1]Augusta Univ., Augusta, GA, [2]Praxisgenomics, Atlanta, GA, [3]Univ. of Iowa Hosp. and Clinics, Iowa, IA
Disclosure Block: **N. Sahajpal:** None.
Current cytogenetics methods, such as FISH and karyotyping are time-consuming, labor-intensive, and low throughput. While chromosomal microarrays (CMAs), commonly used in clinical diagnosis, are fast and high throughput, they cannot detect translocations and inversions. Optical genome mapping (OGM) generated by the Bionano Saphyr platform provides a direct, high-resolution view of intact, long DNA molecules, which are then assembled *de novo* to reconstruct the native structures of chromosomes. OGM can detect germline structural variants (SVs) ranging from 500bp insertions and deletions to complex chromosomal rearrangements. Here, we perform a multi-site, multi-operator, multi-instrument validation study to determine the robustness and sensitivity of OGM in detecting SVs and copy number variants (CNVs) associated with constitutional genetic disorders. In the current pilot phase of this IRB approved study of a total of 30 samples comprising 13 whole blood samples of patients with several clinical features of neurodevelopmental disorders and with pathogenic CNVs detected by clinical CMAs, 7 Coriell cell lines with known CNVs, SVs, *FMR1* repeat expansions, and *DUX4* repeat contractions, and 10 whole blood samples from phenotypically normal human volunteers were tested by two clinical diagnostic laboratories. The operators and analysts were blinded to the known genetic abnormalities; and only received the clinical phenotype information. A standard operating procedure (SOP) was devised to enable analysts to systematically and efficiently select for rare (

PrgmNr 3083 - To report or not to report: The power of curating gene-disease relationships with limited evidence for rare disease

View session detail

Author Block: **A. Clause**[1], J. P. Taylor[1], R. Rajkumar[1], K. Bluske[2], E. Thorpe[1], D. R. Bentley[1], R. J. Taft[1], D. L. Perry[1], A. J. Coffey[1]; [1]Illumina, Inc., San Diego, CA, [2]Illumina, Inc, San Diego, CA
Disclosure Block: **A. Clause:** Salary/Employment; Illumina, Inc.. Other; minor shareholder in Illumina, Inc..
The Illumina Clinical Services Laboratory uses the ClinGen Gene-Disease Validity framework to support reporting for the TruGenomeâ□¢ Undiagnosed Disease clinical whole genome sequencing (cWGS) test for patients with suspected rare and undiagnosed genetic disease. Gene-disease relationships (GDRs) of interest are reactively curated during case analysis. Between Oct. 2018 and Feb. 2021, we curated 292 GDRs to support interpretation and reporting for 638 families. Of these GDRs, 243 (83%) had sufficient evidence to support reporting a variant as a potential diagnostic finding (Moderate, Strong, or Definitive GDR), whereas ten were classified as Disputed or No Known Disease Relationship. Variants in genes that lack an established disease relationship are also of value to report provided there is evidence that the GDR is relevant to the probandâ□□s phenotype and supportive of variant pathogenicity. Such variants can be reported as research candidates, highlight a possibly novel diagnosis, or act as a flag for future reanalysis as new evidence emerges.
Thirty-nine GDRs (13%) were considered of uncertain significance, with 36 classified as having Limited evidence and three as No Known Disease Relationship - Animal Model Only. Variants of uncertain significance were returned in 26 of these genes. Most often, the decision to report was due to compelling overlap in phenotype between the proband and previous cases. For example, variants in the *ENTPD1* gene were reported in a proband with similar features to three unrelated individuals described in the literature. For the remaining 13 GDRs, close scrutiny of the evidence led to a decision not to report a variant due to weak or non-specific overlap between the phenotype of the proband and previously described cases, a mismatch in potential mechanism between the probandâ□□s and reported variants, or because variants in the literature lacked sufficient evidence of pathogenicity. For example, missense variants in the *TNIK* gene were not reported in a proband with a complex phenotype because the lone previous report was a homozygous stop-gained variant in a consanguineous family with non-specific intellectual disability.
Careful evaluation of genes helps to deliver the full promise of cWGS, particularly for individuals with rare disease, whose chance of receiving a potentially informative finding should not be constrained by the absence of large numbers of previously described cases. Gene curation provides a principled means to fully consider all GDRs and allows reporting of variants with valid, albeit limited, evidence of their potential significance.

PrgmNr 3084 - *Kmt5b* is highly expressed in the developing brain and may regulate other known autism risk genes and processes

View session detail

Author Block: **H. Stessman**, R. Wickramasekara, B. Robertson, J. Hallgren; Creighton Univ., Omaha, NE
Disclosure Block: **H. Stessman:** None.
Sequencing of neurodevelopmental disorder (NDD) cases has identified lysine methyl transferase 5B (*KMT5B*) as a high confidence risk gene. We have previously characterized developmental and behavioral phenotypes associated with mouse *Kmt5b* haploinsufficiency revealing social and repetitive grooming phenotypes in females and significant growth and developmental defects in males. KMT5B is an enzyme that di-methylates the lysine 20 residue of histone protein 4 (i.e., H4K20me2). This chromatin mark is thought to function in chromatin compaction/gene silencing and resolution of DNA double stranded breaks; although, its function in the brain is largely unknown. This study aimed to (1) catalog the expression pattern of *Kmt5b* over developmental time and (2) identify genes, pathways, and processes impacted by *Kmt5b* loss. Wild-type mice (C57BL/6N) were time mated to produce embryos and pups for the temporal detection of *Kmt5b* transcripts using RNAScope chromogenic *in situ* hybridization. Whole-mount embryos were collected and fixed at embryonic days (E) E11.5, 12.5, 13.5, 14.5, and 16.5, and whole brains at postnatal days (P) 1-2, 10, and 42. RNAScope revealed strong *Kmt5b* expression across the developing central nervous system that was highest during embryonic development and reduced postnatally. To better understand how KMT5B regulates global gene expression, we performed RNA-sequencing (RNA-seq) at E14.5 on whole brain homogenates. *Kmt5b* haploinsufficiency was modeled using a gene trap mouse ($Kmt5b^{gt}$). HET ($Kmt5b^{+/gt}$) x HET pairs generated litters of WT (*K*mt5b+/+), HET, and KO (*Kmt5b*gt/gt) embryos. Three brains of each genotype (WT, HET, and KO) and sex were collected for analysis. The primary effect of *Kmt5b* loss was gene upregulation; the only gene downregulated across all HET and KO conditions was *Kmt5b*. While only ten differentially expressed genes (DEGs) were identified between WT male and female brains, the impact of *Kmt5b* loss appeared more severe in males (more cell stress and death) than females identified through gene set enrichment analysis (GSEA). These analyses also indicated that DNA damage may be a significant source of cell stress and cell cycle delays in HET and KO brains. Finally, we identified a significant enrichment between our upregulated DEGs and known *CHD8* downregulated DEGs (q=0.025). Of the genes that contributed to this enrichment, many are known to regulate cell population proliferation (Stringdb; GO:0042127; FDR q=1.23e-05). These results are particularly interesting given that *CHD8* haploinsufficiency has been associated with overgrowth, and we have identified undergrowth in our *Kmt5b* mouse model.

PrgmNr 3085 - *TRAPPC10* mutation is associated with a microcephalic TRAPPopathy disorder in humans and mice

View session detail

Author Block: **H. Almousa**[1], L. Rawlins[2], S. Khan[3], S. Collins[4], M. Milev[1], J. Leslie[2], D. Saint-Dic[1], A. Hincapie[1], G. Harlalka[2], V. Vancollie[5], C. Lelliott[5], A. GUL[6], B. Yalcin[4], A. H. Crosby[7], M. Sacher[1], E. L. Baple[8]; [1]Concordia Univ., Montreal, QC, Canada, [2]Univ. of Exeter, Devon, United Kingdom, [3]Intl. Islamic Univ., Islamabad, Pakistan, [4]Univ. of Bourgogne Franche-ComtÃ©, Dijon, France, [5]Wellcome Sanger Inst., Hinxton, United Kingdom, [6]Intl. Islamic Univ., Islamabad, Islamabad, Pakistan, [7]Univ. of Exeter Med. Sch., Exeter, United Kingdom, [8]Univ. of Exeter, Exeter, Exeter, United Kingdom
Disclosure Block: **H. Almousa:** None.
Proteins and other macromolecules are sorted throughout the cell in membrane bound vesicles in a process known as membrane trafficking. Defects in this process are associated with various human disorders. The highly evolutionarily conserved transport protein particle (TRAPP) complexes (TRAPP II and III) perform fundamental roles in subcellular trafficking pathways. This study aims to assess the effect of two homozygous variants in *TRAPPC10*, a component of the TRAPP II complex, in individuals with a microcephalic neurodevelopmental disorder. Here, we present comprehensive genetic, clinical, and functional data to assess the effect of these variants on membrane trafficking and ciliogenesis. Molecular studies revealed a reduced interaction between mutant TRAPPC10 and its putative adaptor protein TRAPPC2L. Studies of patient lymphoblastoid cells revealed an absence of TRAPPC10 as well as an unexpected absence of TRAPPC9, another key TRAPP II complex component associated with a clinically overlapping neurodevelopmental disorder. The TRAPPC9/10 reduction phenotype was recapitulated in TRAPPC10$^{-/-}$ knockout cells, which also revealed a membrane trafficking and ciliogenesis defect. Notably, both the reduction in TRAPPC9 levels and the trafficking defect could be rescued by wild type but not mutant TRAPPC10. Studies of *Trappc10*$^{-/-}$ knockout mice revealed significant neuroanatomical brain defects and microcephaly, paralleling those seen in the human condition as well as in a *Trappc9*$^{-/-}$ mouse model. Together these studies confirm *TRAPPC10* gene mutation as a cause of human disease and define TRAPP-mediated pathomolecular outcomes of importance to TRAPPC9/TRAPPC10 mediated neurodevelopmental disorders in humans and mice.

PrgmNr 3086 - Advanced Diagnostics and Genotype-Phenotype Resolution using Functional Genomics in >500 Neuromuscular and Neurological Disorder Patients

View session detail

Author Block: **S. Chakravorty**[1], K. Berger[2], L. Rufibach[3], S. Shira[3], S. Verma[1], R. Logan[1], M. Wicklund[4], M. B. Harms[5], T. Mozaffar[6], V. E. Kimonis[7], D. Arafat[2], G. C. Gibson[8], M. R. Hegde[9]; [1]Emory Univ., Atlanta, GA, [2]Georgia Inst. of Technology, Atlanta, GA, [3]Jain Fndn. Inc., Seattle, WA, [4]Univ. of Colorado Neurology, Denver, CO, [5]Washington Univ. Sch. of Med., St. Louis, MO, [6]Univ. of California Irvine Neurology, Irvine, CA, [7]Univ CA Irvine, Orange, CA, [8]Georgia Tech, Atlanta, GA, [9]PerkinElmer, Lilburn, GA

Disclosure Block: **S. Chakravorty:** None.

Background: 50-70% of inherited rare neuromuscular disease (NMD) patients remain undiagnosed even after DNA testing, a barrier for clinical trial enrolment. Recently, using a muscular-dystrophy next-generation-sequencing panel on 4656 congenital/limb-girdle muscular dystrophy (CMD/LGMD)-suspected patients across the US and using exome sequencing on 207 genetic myopathies across the Indian subcontinent, and investigating hereditary NMDs and peripheral neuropathies at CHOA, we identified the major hurdles were: a) lack of genotype-phenotype correlation, b) high prevalence (72%) of variants of uncertain significance (VUSs), c) >30% of all patients had pathogenic variant(s) or VUSs in â¥2 genes (multi-genic), and d) the lack of less-invasive biomarker-driven approaches. **Methods:** Here, we used high-throughput clinical-grade RNA-Seq with a proprietary tiered analytical pipeline, and co-immunoprecipitation combined with mass-spectrometry proteomics, and other targeted assays on muscle/skin/blood biopsy-tissues to resolve VUSs and multi-genic cases to enhance molecular diagnostics, and to resolve genotype-phenotype relationships. **Results:** Using targeted RNA-Seq analysis and other assays with genotype-clinical-data correlation on 548 cases, we achieved 64% diagnostic yield and 88% diagnostically informative results. Besides VUS reclassification, we identified variant mechanisms acting either by abnormal splicing/allele/gene-expression/protein-stability/function levels or causing defects in pathways. For example, an 8-year old child with proximal weakness, dystrophic changes on muscle biopsy with normal immunohistochemistry including alpha sarcoglycan stains harbored two variants (pathogenic: c.229C>T, likely pathogenic: c.957-11C>G) *in trans* in *SGCA*. This genotype-muscle biopsy discrepancy was resolved using above-mentioned technique by identifying normal mRNA and protein expression for all Sarcolgycans in muscle even with c.957-11C>G causing both exons 6-7 and 6-8 skips, and that the variantsâ pathogenicity acting at protein function level. Furthermore, the novel application helped discover a new gene, *DRGX*, associated with bilateral hand weakness, finger flexor contractures and sensory motor polyneuropathy in a teenager. Additionally, using enzyme assays and RNA-Seq, we reclassified 20 *GAA* (Î±-glucosidase) gene VUSs as pathogenic variants to resolve undiagnosed Pompe disease cases. **Conclusions:** Our results show the importance of using a multi-tiered approach that includes omics platforms, biomarkers and genotype-phenotype correlation not only for diagnostics but also for better trial-readiness.

PrgmNr 3087 - Cell culture GWAS identifies common genetic variants that influence lithium induced neural progenitor proliferation

View session detail

Author Block: **J. Wolter**[1], B. Le[2], N. Matoba[3], J. Valone[1], A. Elwell[1], N. Aygun[4], D. Liang[5], M. Lafferty[5], K. Courtney[6], J. Piven[7], M. Zylka[1], J. Stein[4]; [1]Univ. of North Carolina, Chapel Hill, NC, [2]UNC Chapel Hill, Durham, NC, [3]UNC at Chapel Hill, Chapel Hill, NC, [4]UNC Chapel Hill, Chapel Hill, NC, [5]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [6]NIH/NIMH, Bethesda, MD, [7]Univ North Carolina, Chapel Hill, NC
Disclosure Block: **J. Wolter:** None.
Lithium is highly effective in the treatment of bipolar disorder, partly through its ability to induce neural progenitor cell (NPC) proliferation and adult neurogenesis. Yet a substantial portion of patients do not respond to lithium, and recent evidence suggests that genetic variation plays a role in lithium sensitivity. Combining genome wide association studies (GWAS) with experimental perturbations in genetically diverse human primary cell lines is an emerging and powerful approach to reveal pharmacogenomic interactions in a tightly controlled setting. Here, we use a library of 84 genetically diverse primary human neural progenitor cell lines to identify genetic variation which modulates lithium induced proliferation. We identified a region on chromosome 3 which modulates proliferation from a therapeutically relevant dose of lithium. This locus correlates with change in proliferation rate, and not baseline proliferation, suggesting a bona fide pharmacogenomic interaction. This locus has previously been implicated in GWAS for bipolar disorder, lithium responsiveness in bipolar patients, and general intelligence. Within this gene dense region, we identified one gene, *GNL3*, which is responsive to lithium treatment. We used a polymorphism in *GNL3* to quantify lithium induced allele specific *GNL3* expression, finding that increased levels of *GNL3* correlated with increased proliferation. Increasing *GNL3* expression with dCas9:VP64 induced NPC proliferation, and decreasing *GNL3* expression with dCas9:KRAB reduced lithiumâ□□s effect on proliferation. Looking forward, these experiments motivate the exploration of predicting bipolar patient responsiveness, and modulating *GNL3* expression to affect lithium responsiveness *in vivo*.

PrgmNr 3088 - Characterization of molecular effects of a novel *GNAQ* mutation identified in Sturge-Weber Syndrome

View session detail

Author Block: **F. Galeffi**[1], D. A. Snellings[2], S. Wetzel-Strong[2], J. Bullock[2], C. Gallione[3], N. Kastelic[1], P. E. North[4], D. A. Marchuk[3]; [1]Duke Univ. Med. Ctr., Durham, NC, [2]Duke Univ., Durham, NC, [3]Duke Univ Med Ctr, Durham, NC, [4]Med. Coll. of Wisconsin, Milwaukee, WI
Disclosure Block: **F. Galeffi:** None.
Sturge-Weber syndrome (SWS) is a sporadic, congenital, neuro-cutaneous disorder characterized by a capillary vascular malformation with or without brain lesions. This syndrome is caused by a somatic-mosaic, activating mutation in *GNAQ* that encodes the G protein subunit alpha-q protein. Although the missense mutation R183Q is the sole *GNAQ* mutation identified thus far in affected tissues of ~ 90% of SWS patients, the most common *GNAQ* mutation in cancer is the constitutively-activating Q209L mutation. The restricted mutation spectrum for SWS and related capillary malformations suggests that Q209L might not be tolerated in vascular development. In this study, we sequenced skin biopsies of affected capillary malformations from an additional 10 SWS patients and as expected, found the R183Q mutation in 90% of SWS patients. However, one sample exhibited a Q209R mutation in 113/681 reads (16.6%), which is a rare variant (1.2% of *GNAQ* mutations) in the COSMIC database. To compare the effect of the Q209R mutation on downstream G protein signaling with other relevant *GNAQ* mutations, we performed Luciferase reporter assays in human embryonic kidney cells (HEK293T) transfected with a *GNAQ*-responsive luciferase reporter co-transfected with either *GNAQ* WT, R183Q, Q209L, Q209R, or C9X (premature termination codon representing a null allele). Expression of Q209R induced a 20-fold increase in luciferase activity compared to either WT or null alleles, while Q209L and R183Q induced increases of 50- and 10-fold, respectively. These results confirmed the activating role of Q209R, which was similar to R183Q, but significantly lower than Q209L (***pGNAQ variants. All of the R183 and Q209 missense variants caused extensive dysregulation of a broad range of transcripts compared to the WT or null allele, confirming that these are all activating mutations. However, the missense variants exhibited few differentially expressed genes (DEGs) when compared to each other. These data suggest that these mutations differ in magnitude of activation, but have similar downstream effects. KEGG analysis of DEGs between WT and Q209L revealed several dysregulated signaling pathways including TNF, IL-17, and NF-kappa B, suggesting an inflammatory response to *GNAQ* activation. Gene set enrichment analysis of these DEGs implicates signaling through EGFR, KRAS, and MTOR, consistent with some of the known downstream effectors of *GNAQ* signaling.

PrgmNr 3089 - Clinical and biochemical characteristics of MOGS-CDG, a rare congenital disorder of glycosylation

View session detail

Author Block: **S. Shimada**[1], B. G. Ng[2], A. L. White[3], K. K. Nickander[3], C. Turgeon[3], K. L. Liedtke[4], C. T. Lam[5], E. Font-Montgomery[6], C. M. LourenÃ§o[7], M. He[8], L. A. UmaÃ±a[9], C. L. Uhles[10], P. Wheeler[11], M. J. Bamshad[12], D. A. Nickerson[12], R. Gates[13], H. M. Byers[14], R. Sachdev[15], T. Cushing[16], L. Smith[17], A. Poduri[18], S. Malone[19], R. Harris[20], I. E. Scheffer[20], D. R. Adams[21], W. A. Gahl[21], M. Malicdan[21], K. Raymond[22], H. H. Freeze[2], L. A. Wolfe[21]; [1]Med. Genetics Branch, NHGRI, NIH, North Bethesda, MD, [2]Human Genetics Program, Sanford Burnham Prebys Med. Discovery Inst., La Jolla, CA, [3]Dept. of Lab. Med. and Pathology, Mayo Clinic Coll. of Med., Rochester, ME, [4]Dept. of Lab. Med. and Pathology, Mayo Clinic Coll. of Med., Rochester, MD, [5]Dept. of Pediatrics, Seattle Childrenâ□□s Hosp. and Univ. of Washington, Seattle, WA, [6]Univ. of Missouri Hlth.Care, Columbia, MO, [7]Faculdade de Med., Centro Univ.rio EstÃicio, RibeirÃ£o Preto, Brazil, [8]Dept. of Pathology and Lab. Med., Childrenâ□□s Hosp. of Philadelphia, Philadelphia, PA, [9]Dept. of Pediatrics, Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [10]Dept. of Genetics, Childrenâ□□s Med. Ctr. Dallas, Dallas, TX, [11]Arnold Palmer Hosp. Pediatric Genetics, Orlando, FL, [12]Dept. of Genome Sci., Univ. of Washington, Seattle, WA, [13]Div. of Med. Genetics, Stanford Univ., Stanford, CA, [14]Div. of Med. Genetics, Stanford Univ., Stanford, Stanford, CA, [15]Ctr. for Clinical Genetics Sydney Childrenâ□□s Hosp. Randwick, Sydney, Australia, [16]Dept. of Pediatrics, Sch. of Med., Univ. of New Mexico, Albuquerque, NM, [17]Dept. of Neurology, Boston Childrenâ□□s Hosp. and Harvard Med. Sch., Boston, MA, [18]Dept. of Neurology, Boston Childrenâ□□s Hosp. and Harvard Med. Sch., Boston, Boston, MA, [19]Dept. of NeuroSci.s, Queensland Childrenâ□□s Hosp., Brisbane, Australia, [20]Austin Hlth.and Royal Children's Hosp., Florey and Murdoch Children's Res. Inst.s, Melbourne, Australia, [21]Med. Genetics Branch, NHGRI, NIH, Bethesda, MD, [22]Dept. of Lab. Med. and Pathology, Mayo Clinic Coll. of Med., Rochester, MN

Disclosure Block: **S. Shimada:** None.

Background: Congenital disorders of glycosylation (CDG) are an increasingly recognized group of disorders due to inborn errors of glycosylation. Over 150 types of CDG due to impaired lipid or defective protein glycosylation have been reported, but precise diagnosis can be challenging because of clinical heterogeneity. Loss of MOGS activity does not block all subsequent *N*-glycan processing, as a Golgi-localized endo-Î±-mannosidase can cleave a tetrasaccharide (Glc_3Man) from the glucosylated branch to allow continued glycan processing. While the majority of CDGs can be identified through carbohydrate deficient transferrin (CDT) testing in blood, some CDGs, including MOGS-CDG, will not be reliably detected using this methodology. We aimed to describes the clinical heterogeneity of this multisystem disorder, doubling the total number report to gain insights in twelve cases of MOGS-CDG, combining biochemical and molecular testing to diagnose. **Methods:** Individuals with biallelic variants in *MOGS* were identified by exome sequencing and targeted arrays through an international, multicenter collaboration. Phenotypes and associated biochemical findings were assessed to determine the pathogenicity of individualsâ□□ genetic variants. Confirmatory biochemical assays were performed on serum and urine. **Results:** Biallelic variants in *MOGS* were identified in 12 individuals from 11 families, including 10 unreported individuals. The clinical heterogeneity in MOGS-CDG was delineated, and the neurologic, immunologic, and skeletal phenotypes were described in detail. Urine oligosaccharide analysis was consistently abnormal for all affected probands, whereas biochemical analysis of serum glycosylation was not consistent. **Conclusions:** The clinical phenotypes of MOGS-CDG include multisystemic involvement with variable severity. Molecular analysis, combined with biochemical testing, is important for diagnosis. In MOGS-CDG, urine oligosaccharide analysis can be used as a reliable biochemical test for screening and confirmation of disease.

PrgmNr 3090 - CRISPR engineering of a dystonia-specific allelic series of *THAP1* mutations reveals transcriptional profiles associated with myelination and neurodevelopment

View session detail

Author Block: **A. Domingo**[1], R. Yadav[2], S. Shah[1], W. T. Hendriks[1], S. Erdin[2], D. Gao[2], K. O'Keefe[1], B. Currall[1], J. F. Gusella[2], N. Sharma[2], L. J. Ozelius[2], M. E. Ehrlich[3], C. Bragg[1], M. E. Talkowski[2]; [1]Massachussetts Gen. Hosp., Boston, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Icahn Sch. of Med. at Mt. Sinai, New York, NY

Disclosure Block: **A. Domingo:** None.

The dystonias are a group of neurologic disorders characterized by disturbances in volitional movement. Its genetic architecture is heterogenous, and genetic forms are associated with diverse variants in multiple genes; whether there are shared molecular pathways across mutations/genes remains uncertain. Here we integrated disease modeling using CRISPR-modified induced pluripotent stem cells (iPSCs) with transcriptomic analyses to investigate the molecular consequences of variation in the gene *THAP1*, where dystonia-associated mutations that localize to specific domains are thought to alter gene function differently. We engineered an allelic series of eight variants in a common iPSC background and differentiated these lines into a panel of near-isogenic neural stem cells (n = 94 lines). The mutant lines thus harbored one of pathogenic N12K, S21Thet/hom, P26R or C54Y mutations in the DNA-binding domain, or an R146fs in the nuclear localization sequence (NLS), an M143V variant predicted to be benign, or a deletion of the entire *THAP1* coding region. These were compared to CRISPR-targeted unedited cells and untargeted iPSC lines to discover common differentially expressed genes (DEGs). Each mutation model induced significant (FDR9.6-fold enrichment for shared DEGs compared to null expectation models. Given this, we derived a joint DEG list using weighted Z-scores to define the 871 genes consistently altered by mutations in these domains (FDRThap1-disruptive alleles (C54Y and exon-2 deletion) and detected significant changes in myelin gene expression on targeted expression assays and reduction of myelin structural integrity using staining (PTHAP1 mutations.

PrgmNr 3091 - Distribution and allele frequency of the*Gln3060X*variant present in the*ASPM*gene associated to microcephaly and related disorders in Puerto Rico

View session detail

Author Block: **J. Ã□vila PagÃin**; Univ. of Puerto Rico: MayagÃ¼ez Campus, San SebastiÃin, Puerto Rico
Disclosure Block: **J. Ã□vila PagÃin:** None.
Distribution and allele frequency of the *Gln3060X* variant present in the *ASPM* gene associated to microcephaly and related disorders in Puerto Rico. Microcephaly is a disorder that affects brain-development, resulting in neonates with a smaller head-circumference than expected, generally 3-4 SD below the mean depending on age and gender. The main focus of this research project is the condition Primary Hereditary Microcephaly (MCPH), microcephaly discovered before birth that is commonly caused by genetic factors. Although many genes have been identified as risk factors for MCPH, mutations in the *ASPM* gene are known to be the biggest risk factor associated with primary microcephaly. The *ASPM* gene is known to play a very important role in mitotic division during neurogenesis in a developing brain. Therefore, a mutation in this gene could be the cause of Microcephaly or conditions commonly related to it. The particular variant being studied is *Gln3060X*, a missense variant of the *ASPM* gene previously associated with primary microcephaly. A general population study in Puerto Rican population was done using an array of 625 DNA samples from healthy individuals evenly divided by 30 municipalities around Puerto Rico. The variant was tested using the ViiA 7 Real Time PCR system, using TaqMan and the particular SNP (rs137852994) for the *Gln3060X* variant. Among these 625 samples, 62 mutated individuals were found; 61 heterocigotes (G/A) and 1 homocigote (A/A). In total, this study found an approximate 5% allelic frequency of the *Gln3060X* mutation in Puerto Rican population. Therefore, we can confirm the presence and prevalence of this mutation in Puerto Rico, mainly concentrated among North and East regions of the Island. Furthering the project, patients that suffer from Microcephaly and related conditions will be collected to participate in the study to prove allelic frequency in patients that suffer from them, meanwhile associating this variant with the development of Microcephaly and other conditions in Puerto Rico.

PrgmNr 3092 - Enhancement of CHRNA5 mRNA expression by upstream polymorphisms associated with preserved cognition and lessened neuropathology in an aged human population

View session detail

Author Block: **J. Rybnicek**[1], Y. Chen[2], D. A. Bennett[3], P. L. De Jager[4], H-U. Klein[5], S. Tripathy[2], D. Felsky[2], E. K. Lambe[1]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [3]Rush Univ., Chicago, IL, [4]Columbia Univ Med Ctr, New York, NY, [5]Columbia Univ., New York, NY

Disclosure Block: **J. Rybnicek:** None.

Single nucleotide polymorphisms (SNPs) in cholinergic system genes have been previously linked to Alzheimerâ□□s disease (AD). However, the genetic and cellular mechanisms of these associations remain unexplored. The identification of SNP effects at the level of gene expression and AD-related neuropathologies in human brain could provide evidence for the stratification of AD subjects into molecular subtypes and inform novel treatment strategies. We analyzed *ante-mortem* cognition and *post-mortem* neuropathology data from 1,050 elderly human subjects from the Religious Orders Study and Memory and Aging Project cohorts. All subjects had matched genotypic and *post-mortem* RNA sequencing data from prefrontal cortex (PFC), and 24 had PFC single-nucleus data available. Using general linear models, we replicated a previously-described positive effect of a six-SNP haplotype in an upstream region of *CHRNA5* on levels of *CHRNA5* mRNA in PFC bulk-tissue and single-nucleus data. *CHRNA5* codes for the Î±5 subunit of the nicotinic receptor, which has been previously linked to nicotine dependence and attention deficits, but not AD. Upon further investigation, we identified a novel, sex-dependent association of this regulatory *CHNRA5* haplotype with preserved cognitive status at death as well as lower levels of beta-amyloid and phosphorylated tau in brain. Together, our findings suggest a neuroprotective role for elevated *CHRNA5* expression in brain, and ongoing work is investigating closely-associated cellular and molecular pathways.

PrgmNr 3093 - Functional characterization of haplotype surrounding *TOMM40*-523â repeat to assess differential risk effects on European ancestry *APOEÉ3* haplotypes

View session detail

Author Block: **M. Lipkin Vasquez**, P. Bussies, F. RAJABLI, K. L. Hamilton-Nelson, A. J. Griswold, M. A. Pericak-Vance, J. Young, J. M. Vance, K. Nuytemans; John P. Hussman Inst. for Human Genomics, Miami, FL
Disclosure Block: **M. Lipkin Vasquez:** None.
Introduction: Reports on involvement of *TOMM40*, a gene neighboring *APOE*, in Alzheimer Disease (AD) risk have been inconsistent. Recently, we showed that the length of a poly-T repeat in *TOMM40* (*TOMM40*-523â) is associated with AD risk in individuals carrying *APOEÉ3*, the most common *APOE* haplotype in the general population, on European local ancestry (LA). âVery longâ repeat lengths (VL, >29T) have a protective effect compared to âshortâ repeat lengths (S, APOE*É3* or *APOEÉ4* haplotypes in either ancestral background. Therefore, we hypothesized that variants in linkage disequilibrium (LD) with *TOMM40*-523â on the European LA *APOEÉ3* haplotype can modify risk for AD, potentially through *APOE* regulation. Methods: We used the short tandem repeat detection bioinformatics algorithm HipSTR to type S and VL repeats in whole genome sequencing data of individuals homozygous for the *APOEÉ3* European LA haplotype from the Puerto Rico AD Initiative (PRADI) project. Frequency of variants on 16 S and 14 VL independent haplotypes were compared to determine variants in LD with the repeat. HaploView was used to determine the LD structure of the repeat and surrounding variants. Results: We identified a 16kb LD block surrounding *TOMM40*-523â harboring 21 variants in strong LD (r^2>0.9) with the repeat (hg19, chr19:45,395-45,411k). This region includes the putative *APOE* promoter (harboring LD variants rs405509, rs449446 and rs769450) and a *TOMM40* intronic region with previously reported enhancer activity (harboring LD variants rs157580, rs2075649 and rs157584). Assessment of combined regulatory function of variants in LD with the repeat on S or VL haplotypes in the *APOE* promoter and *TOMM40* enhancer region is currently ongoing using luciferase reporter assays in AD-relevant cell types (i.e. astrocytes, microglia and neurons). Discussion: The identification of clearly distinct S and VL haplotypes on *APOEÉ3* European LA background support importance of the surrounding variants in the risk effect observed in the association analyses. The LD analyses data suggest that *TOMM40*-523â LD variants could directly affect *APOE* expression through presence in the promoter itself and/or in an identified enhancer region in *TOMM40*. The follow-up functional data will pinpoint the driving regulatory element(s) and *TOMM40*-523â LD variants for the observed different AD risk effects in European ancestry *APOEÉ3* carriers. Long term, treatments targeting these regulatory regions may be relevant to a large amount of people given *APOEÉ3*âs frequency in the general population.

PrgmNr 3094 - Identification of candidate genes for alcohol use disorder using RNA-seq data

View session detail

Author Block: **A. Q. Nato**[1], H. Ata Ul Mustafa[1], H. G. Sexton[1,2], S. D. Moore[3,4], J. Denvir[1], D. A. Primerano[1], M-L. Risher[1,2,3]; [1]Dept. of BioMed. Sci., Joan C. Edwards Sch. of Med., Marshall Univ., Huntington, WV, [2]Hershel â□□Woodyâ□□ Williams Veteran Affairs Med. Ctr., Huntington, WV, [3]Dept. of Psychiatry and Behavioral Sci., Duke Univ. Med. Ctr., Durham, NC, [4]Durham Veteran Affairs Med. Ctr., Durham, NC

Disclosure Block: **A.Q. Nato:** None.

Alcohol use disorder (AUD) is a chronic psychiatric disorder with complex etiology. Individuals with AUD usually have intense drinking patterns that result in acute and long-term effects on the brain. Binge drinking results in high alcohol intake, which is common among adolescents and young adults, and is associated with increased risk of developing AUD. Here, we analyze RNA-seq data from hippocampus of Sprague Dawley rats (*Rattus norvegicus*) to investigate temporal variations in gene expression in the presence or absence of ethanol. We used a rodent model of binge drinking (adolescent intermittent ethanol (AIE)) to identify candidate genes that may contribute to the long-term effects on brain function and AUD development. At postnatal day (PND) 30 (adolescence), rats received chronic intermittent ethanol (5g/kg intragastrically (i.g.) 10 times across 16 days). RNA extracted from hippocampal tissue was collected at three time points: 1) 24 hours after 4th dose (PND35), 2) 24 hours after last dose (PND46), and 3) 24 days after last dose (PND70; adult), and sequenced. We processed RNA-seq data (TrimGalore), compiled gene counts (HTSeq), and performed differential expression analysis (DESeq2). We performed fast gene set enrichment analysis (fgsea) using the effect size estimate (log2 fold change) and gene sets of Reactome and KEGG pathways accessed through these databases and MSigDB. We also employed gene graph enrichment analysis (GGEA) using the same data. We visualized the results using the Cytoscape pathway analysis package and identified subnetworks of genes enriched by higher interaction, through the Cytoscape clusterMaker app, which implements Markov CLustering Algorithm (MCL). Using these results, we focused on genes with particular functions, cell types, or diseases such as addiction, neuronal/synaptic remodeling, immune response, inflammation, blood brain barrier, and aging. Candidate genes determined from these analyses may provide clues to the underlying mechanism for detrimental effects of AIE exposure that contribute to the development of AUD.

PrgmNr 3095 - Influence of *APOE e4*in neuropathologic lesions of Alzheimer disease and related dementias

View session detail

Author Block: **D. Godrich**[1], J. Pasteris[1], E. R. Martin[2], G. D. Schellenberg[3], M. A. Pericak-Vance[4], M. L. Cuccaro[5], W. Kukull[6], T. Montine[7], G. W. Beecham[1], The Alzheimer's Disease Genetics Consortium; [1]Univ. of Miami, Miami, FL, [2]Univ. of Miami Miller Sch. of Med., Miami, FL, [3]Univ Pennsylvania Sch Med, Philadelphia, PA, [4]Miami, FL, [5]Univ Miami Sch Med, Miami, FL, [6]Univ. of Washington, Seattle, WA, [7]Stanford Univ., Palo Alto, CA

Disclosure Block: **D. Godrich:** None.

Background. Alzheimer disease (AD) is a highly heritable disease that is partially driven by genetics, though much of it is still unexplained. The strongest known genetic risk factor for AD is the *APOE* Îµ4 allele. However, this risk has mostly been shown with the endpoint of clinically diagnosed AD. To get closer to the underlying genetics, it is important to investigate the pathological changes seen at autopsy in the AD brain including neurofibrillary tangles (NFT), neuritic plaques (NP), and diffuse plaques (DP). Moreover, it is uncommon for AD lesions to occur in isolation; instead, they often appear comorbidly with AD related dementia (ADRD) lesions including Lewy Bodies (LB), cerebral amyloid angiopathy (CAA), arteriolosclerosis (ARTE), hippocampal sclerosis (HS), vascular brain injury (VBI), and TAR DNA-binding protein 43 (TDP-43) inclusions. Uncovering genetic factors underlying AD/ADRD lesions using genome-wide association studies (GWAS) may help fill in the gaps of the unexplained genetics. Herein, we report on the influence of the *APOE* Îµ4 status on different AD/ADRD lesions as a first step in the investigation of the genetic architecture of AD/ADRD neuropathologic lesions.

Methods. We used the National Alzheimerâ□□s Coordinating Center (NACC) database, identifying 4,972 autopsied individuals with a neuropathology assessment and *APOE* Îµ4 genotyping. Lesions were ranked as ordinal endpoints based on increasing severity. *APOE* status was categorized as 0, 1, or 2 copies of the *APOE* Îµ4 allele. Correlation between lesions and *APOE* Îµ4 was assessed. Associations between *APOE* Îµ4 and neuropathologic lesions were analyzed using ordinal logistic regression models adjusting for age at death and sex.

Results. *APOE* Îµ4 showed moderate correlations to the three AD lesions and CAA, and minor negative correlations with ARTE and VBI. Ordinal logistic regression models showed that carrying one or more copies of *APOE* Îµ4 is associated with presence of more severe neuropathologic lesions of NFT (OR = 2.13, p-value**Conclusions.** *APOE* Îµ4 is associated to severity of AD lesions, CAA, LB, and TDP-43, but not associated with severity of other ADRD lesions like ARTE, HS, and VBI. These data show that not all ADRD lesions are associated with *APOE* Îµ4 and suggest that there is a need to search for genetic factors underlying ADRD lesions. To expand on these preliminary findings, we are conducting a GWAS on AD/ADRD neuropathologic lesion endpoints to search beyond *APOE* Îµ4 for novel genetic factors.

PrgmNr 3096 - Molecular mechanisms of autism-associated variants in KCNQ3 channel

View session detail

Author Block: **D. M. Dykxhoorn**, A. Hinojo-Perez, J. Rogers, M. A. Edmond, R. Barro-Soria; Univ. of Miami, Miami, FL
Disclosure Block: **D.M. Dykxhoorn:** None.
Background: Autism spectrum disorders (ASDs) comprise a constellation of clinically and genetically heterogeneous neurodevelopmental disorders that affect 1 in 54 children in the US and are characterized by communication deficits, restrictive social interactions, and repetitive behaviors. Genetic analyses have identified loss- or gain-of-function variants in genes coding for several ion channels, including voltage-gated potassium (Kv) channels. Kv channels are widely expressed in the central nervous system (CNS) where they set the resting membrane potential, shape the duration of action potentials, and regulate neuronal firing. Kv channel dysfunction is a major factor in excitatory/inhibitory (E/I) disequilibrium and network impairment. Specifically, recent studies have identified ASD-associated variants in the voltage sensor (S4) of Kv7.3 channel (R227Q, R230S, R230H, R230C, R230L, and R236C) encoded by the KCNQ3 gene.**Methods:** We use voltage clamp fluorometry (VCF) to measure both S4 and gate movement in wild type and ASD-associated Kv7.3 variants expressed in Xenopus oocytes to determine how mutations affect the movement of these domains in Kv7.3 channels. We are also studying the functionality of the Kv7.3 channel in induced pluripotent stem cell (iPSC)-derived inhibitory neurons bearing these variants.**Results:** We found that the R230C mutation in the S4 of Kv7.3 channels caused voltage-independent channels by shifting the S4 movement to strong negative voltages such that the channel remains open at physiological voltages. Similarly, the mutations R230S and R230L removed voltage-independent gating, resulting in enhanced channel function compared to wt Kv7.3 channels. In contrast, R236C shifted both the channel opening and S4 movement towards positive voltages compared to wt Kv7.3 channels. In R227Q, R230C, R230H, and R236C the time course of the fluorescence and channel re-opening follow each other. However, while R230H had faster activation time course, R236C had a slower activation time course compared to wt Kv7.3 channels. Constructs have been derived for the ASD-associated variants and transduced into control iPSC-derived cortical neurons that are being evaluated for their electrophysiological properties.**Conclusions:** Although these variants have been associated with autism, they alter the functionality of the Kv7.3 channel in different ways. The variants (R227Q, R230C, and R230H) and R236C shift the channel opening and S4 movement in opposite direction, thereby causing a gain- and loss-of Kv7.3 channel function. The results of these studies suggest that these ASD-associated variants might alter neuronal network function.

PrgmNr 3097 - Multi-region single-cell dissection of brain vasculature changes in Alzheimerâ□□s Disease

View session detail

Author Block: **N. Sun**[1,2], L. Akay[1,3], M. Murdock[1,3], Y. Park[4], H. Mathys[1,3], K. Galani[1], L-L. Ho[1], D. Bennett[5], L-H. Tsai[1,3], M. Kellis[6]; [1]Massachusetts Inst. of Technology, Cambridge, MA, [2]Broad Inst., Cambridge, MA, [3]The Picower Inst. for Learning and Memory, Cambridge, MA, [4]Univ. of British Columbia, Vancouver, BC, Canada, [5]Rush Univ., Chicago, IL, [6]MIT & Harvard, Cambridge, MA
Disclosure Block: **N. Sun:** None.
The blood-brain-barrier (BBB) is crucial to maintaining the environment of brain homeostasis through avoiding the entrance of pathogens and hazardous substances in the circulating blood into the brain and supplying neuronal and glial cells with oxygen, nutrients and energy molecules. Impaired BBB function is associated with multiple neurodegenerative diseases, including Alzheimerâ□□s disease (AD). The vascular hypothesis of AD thought the brain vascular damage is the initial event to cause the BBB dysfunction and eventually neuronal injury and amyloid-beta (Ab) accumulation in the brain. However, human cerebrovascular cells remain poorly characterized due to their sparsity and dispersion, and their transcriptional changes in response to neurodegenerative diseases and disease predisposition remain unknown. Here, we report a comprehensive single-cell characterization of the human cerebrovasculature using post-mortem in-silico sorting of human brain tissues. We capture 22,514 cerebrovascular cells across 11 subtypes, including three subtypes of endothelial cells, two subtypes of pericytes, two subtypes of smooth muscle cells, three distinct subtypes of perivascular fibroblasts and ependymal cells, across 440 control and AD individuals, and across 6 brain regions. We determine previously-uncharacterized cell type-specific markers and detect brain-region specific gene expression patterns. We identify 2,676 differentially-expressed genes in total (306 on average) between control and AD individuals at sub-cell-type resolution. Among the top 70 most differentially-expressed genes, we find several genes directly linked to AD-associated genetic variants through both physical and correlation-based enhancer-gene links, and through both tissue-level and brain vasculature-specific eQTLs. These include well-studied AD risk-locus genes ABCB1, SLC14A2, PICALM, which we hypothesize play causal roles in AD through the brain vasculature. We also predict the communications between cerebrovascular cell types and vascular/neuronal/glial cell types, by combining covariation information across individuals, ligand-receptor pairs, and functional association data. Notable examples include dysregulation of the interaction between capillary endothelial cells and pericytes, mediated by DLL4-NOTCH3 signaling, which suggests dysregulation of pericyte survival as a molecular underpinning of AD. Overall, our biological insights and our comprehensive resource molecular atlas of the human cerebrovasculature can help guide future biological and therapeutic studies for AD and other brain disorders.

PrgmNr 3098 - Patterns of gene expression variation in large families carrying developmental delay-associated 16p12.1 deletion

View session detail

Author Block: **A. Tyryshkina**[1], M. C. Jensen[2], L. Pizzo[3], C. Smolen[4], M. Das[5], E. Huber[1], A. Krishnan[6], S. Girirajan[7]; [1]The Pennsylvania State Univ., State College, PA, [2]Pennsylvania State Univ, University Park, PA, [3]University Park, PA, [4]Pennsylvania State Univ., State College, PA, [5]The Pennsylvania State Univ., University Park, PA, [6]Michigan State Univ., East Lansing, MI, [7]Pennsylvania State Univ., University Park, PA

Disclosure Block: **A. Tyryshkina:** None.

The 16p12.1 deletion is associated with complex neurodevelopmental disorders, but the presence of the deletion does not adequately explain the phenotypic variability seen in deletion carriers. Previous studies suggest that additional rare variants in the genome may contribute to phenotypes seen in individuals that carry the deletion. However, the mechanisms of how these rare variants contribute to phenotypic variability remain unknown. To investigate this, we performed whole-genome sequencing, RNA sequencing of lymphoblastoid cell lines, and deep clinical phenotyping on 32 individuals in five large families that carry the 16p12.1 deletion. We found that 1,569 transcripts were differentially expressed between deletion carriers and non-carriers, including the brain development genes *FOXP1*, *ANK3*, and *MEF2*. Differentially expressed transcripts were enriched for genes that are preferentially expressed in the late fetal thalamus (FDR=1.18Ã□10^{-3}), hippocampus (FDR=1.73Ã□10^{-3}), striatum (FDR=0.036), and ventrolateral frontal cortex (FDR=8.21Ã□10^{-3}). We observed additional expression changes in severely-affected children that were either unique or overlapped with their parents, which were enriched for biological functions that matched with 33/41 observed developmental phenotypes in the children. We next tested for enrichment of 25 classes of rare variants towards genes with expression changes, such as outlier expression or alternative splicing. For example, five classes of variants were more likely to be located near genes with outlier expression, including loss-of-function and splice-site variants (FDRZEB2, a gene associated with speech impairment, had elevated expression of *ZEB2* compared to the rest of the cohort, and both siblings presented with speech delay. Finally, we found that genes with â□□second-hitâ□□ variants were more closely connected to genes with expression changes in a brain-specific interaction network compared with permuted networks (p=4.88Ã□10^{-4}). These results suggest that the 16p12.1 deletion and other rare variants in the genome jointly contribute to developmental phenotypes by altering gene expression patterns.

PrgmNr 3099 - Plasma membrane injury may trigger DUX4 expression in FSHD

View session detail

Author Block: **A. Bittel**, Y-W. Chen; Children's Natl. Hosp., Washington, DC
Disclosure Block: **A. Bittel:** None.
Background: Facioscapulohumeral muscular dystrophy (FSHD) is an autosomal dominant neuromuscular disorder caused by the pathological expression of the double homeoprotein 4 (*DUX4*) transcription factor in mature skeletal muscle fibers. *DUX4* activates a number of germline genes and alters RNA and protein metabolism, leading to increase immune responses, apoptosis, and progressive myopathy. We previously demonstrated that immortalized myoblasts from individuals with FSHD demonstrate significant impairments in plasma membrane repair when compared to their unaffected siblings. Rapid repair of the plasma membrane injury is essential for muscle health and involves the activation of cellular pathways that have been shown to contribute *DUX4*expression. However, it is unclear if plasma membrane injury itself triggers *DUX4* expression. **Objective:**To determine if plasma membrane injury leads to *DUX4*expression. If so, to determine the time-course of DUX4 expression during the recovery from injury. **Approach:** We utilized a scrape injury approach to induce membrane injury in thousands of cells simultaneously. Briefly, immortalized myobalsts isolated from the biceps of an individual with FSHD and their healthy sibling were scraped with a cell scraper while incubated in cell-impermeant FITC-Dextran dye. This was followed by a brief incubation in propidium iodide (PI), which binds to double stranded DNA, but is excluded from cells with intact plasma membranes. Using confocal microscopy, we identified cells that were injured and successfully repaired their membranes (Dextran (+), PI(-)), and cells that were injured and failed to repair (Dextran(+), PI(+)). We quantified the percentage of Dextran(+) cells that are injured as a measure of injury susceptibility. Percentage of PI(+) cells was used as marker of recovery capacity. qRT-PCR was used to assess for *DUX4*expression, and expression of its downstream target *ZSCAN4,* 6- and24-hours post-injury. **Results:**Confirming our previous findings, FSHD myoblasts were more susceptible to membrane injury - evidenced by the greater percentage of cells that failed to repair (p<.05 however we did not observe any difference in the percentage of cells that failed to repair pi between fshd and healthy cell lines. also observed significantly increased>DUX4expression at 24 hours post-injury (p<.05 and elevated non-significant>ZSCAN4expression at the same timepoint in FSHD myoblasts. **Conclusion:**The results suggest that membrane injury may contribute to *DUX4* expression and could therefore be an important mechanism of disease progression.

PrgmNr 3100 - Re-analysis of regulatory non-coding promoter/enhancer regions of known congenital myasthenic syndrome genes to identify disease-causing variants

View session detail

Author Block: **K. Polavarapu**[1], R. Thompson[2], L. Matalonga[3], A. Topf[4], S. Laurie[5], Solve-RD DITF-EURO-NMD, SNV-indel WG, J. Ribeiro[6], A. Nalini[7], H. Lochmuller[1,8]; [1]CHEO Res. Inst., Ottawa, ON, Canada, [2]Children's Hosp. of Eastern Ontario Res. Inst., Ottawa, ON, Canada, [3]CNAG-CRG, Barcelona, Spain, [4]John Walton Muscular Dystrophy Res. Ctr., Translational and Clinical Res. Inst., Newcastle Univ. and Newcastle Hosp. NHS Fndn. Trust, Newcastle upon Tyne, United Kingdom, [5]Centro Natl. de AnÃ¡lisis GenÃ³mico (CNAG-CRG), Barcelona, Spain, [6]Neuromuscular Disease Unit, Coimbra Univ. and Hosp. Ctr., Coimbra, Portugal, [7]Dept. of Neurology, Natl. Inst. of Mental Hlth.and NeuroSci.s, Bangalore, India, [8]CNAG-CRG, Ctr. for Genomic Regulation (CRG), Barcelona, Spain
Disclosure Block: **K. Polavarapu:** None.
Promoter and enhancer non-coding regions are among the primary regulatory elements which contain crucial transcription factor binding sites (TFBS). These regulatory regions play a key role in the regulation of gene expression at various levels of transcription and translation. Through targeted sequencing and functional analysis we have previously identified a small number of patients with congenital myasthenic syndrome (CMS) harboring disease-causing variants in the promoter region N-box or in micro-RNA binding site of the 3â□□-UTR. In this study we aimed to analyze these regulatory regions in known CMS genes to identify potential disease-causing variants not previously identified by exome and genome analysis in unsolved cases. Using the GeneHancer database we identified 40 promoter/enhancer regions in 29 known CMS genes. Regions were selected based on gene association score, distance from transcription start site and regions overlapping TFBS sequences with previously reported mutations or having functional evidence. Analysis of regulatory regions was performed in 747 unsolved Neuromuscular disease patients as part of programmatic re-analysis through Solve-RD project. Rare variants with gnomAD and internal MAF A in the regulatory promoter region of the *VAMP1* gene having multiple *RBFOX1* binding sites and another reported frameshift mutation. In another patient with childhood onset ocular onset CMS, we identified likely compound heterozygous variants in *CHRNE* with a nonsense variant and another intronic variant c.500+40G>C affecting a *STAT3* binding regulatory region. In two other suspected CMS patients, regulatory region variants in rare CMS-associated genes *LAMB2* and *PLEC* were identified with partial phenotype correlation requiring further analysis and segregation. This approach of including relevant GeneHancer promoter/enhancer regions might be a relevant addition to the analysis pipeline when no or only single-hit significant coding region variants are detected.

PrgmNr 3101 - Sex bias in neurodevelopmental defects due to differential effects of gene dosage and genetic interactions

View session detail

Author Block: **S. Girirajan**[1], N. Gong[2], A. Sarker[1], B. Scanlon[1], T. Yusuff[1], L. Pizzo[3], M. C. Jensen[4], A. Tyryshkina[5], M. Singh[6], J. S. Iyer[7], M. Das[8], V. Pounraja[9], M. Kayser[2], P. Ingraham[1]; [1]Pennsylvania State Univ., University Park, PA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]University Park, PA, [4]Pennsylvania State Univ, University Park, PA, [5]State College, PA, [6]Natl. Brain Res. Ctr., Gurgaon, India, [7]THE PENNSYLVANIA STATE Univ., UNIVERSITY PARK, PA, [8]The Pennsylvania State Univ., University Park, PA, [9]The Pennsylvania State Univ., State College, PA

Disclosure Block: **S. Girirajan:** None.

Neurodevelopmental disorders have significant sex-bias skewed towards males and are influenced by comorbid features, indicating a complex genetic interplay towards sex-biased phenotypic risk and disease trajectory. Using RNA interference and the UAS-GAL4 system in *Drosophila melanogaster*, we tested the effects of reduced gene dosage and genetic interactions towards differential neuronal, developmental, and behavioral outcomes in males and females. We performed >1100 crosses to test homologs of 136 human genes, including 75 autism and X-linked intellectual disability genes, 32 genes within the neurodevelopmental-associated 3q29, 16p11.2, and 16p12.1 deletions, 29 transcriptional targets of 16p11.2 genes, and 570 pairwise knockdown models using tissue- and neuronal cell type-specific drivers and background specific controls. We used quantitative assays to test sex-specific effects of individual or pairwise knockdown of genes, including neuronal and cellular defects in the fly eye (using *GMR-GAL4* driver), lethality, lifespan, motor defects, seizure susceptibility, and sleep/circadian defects (using *Elav-GAL4* and *nSyb-GAL4* neuronal drivers). We then fine mapped significant sex-biased nervous system defects to distinct neuronal subtypes, including neuroblasts and dopaminergic and glutamatergic neurons. For 6/136 genes (4%), including homologs of *CHD8*, *FMR1*, and *CTNNB1*, more severe developmental and neuronal phenotypes were observed in males. We further tested for interactions of homologs of CNV genes with each other and with genes in neurodevelopmental pathways to identify interactions with biased effects on severity in males and females. Using a multiplicative model, we found additive (18/301, 6%) and synergistic effects (121/301, 40%) of pairwise knockdowns leading to sex-biased severity. For example, male-biased effects were identified with pairwise knockdown of homologs of *KCTD13* and *GRIA3*, while only knockdown of *GRIA3* showed severity in females. Similarly, independently knocking down *PPP4C* and *UBE2A* did not confer any bias in severity, but a strong male-biased effect was observed with combined knockdown of both genes. The sex-biased fly genes have multiple homologs in humans compared to genes with no sex-bias (p

PrgmNr 3102 - Spontaneous DNA damage response, myotonic dystrophy type 1, and CTG tract size

View session detail

Author Block: **V. Li**[1], T. Gall-Duncan[1], S. Lanni[1], L. I. Brady[2], C. Gagnon[3], M. Tarnopolsky[2], J. Mathieu[4], Z. Musova[5], D. Chitayat[6], C. E. Pearson[1]; [1]Genetics and Genome Biology, Hosp. for Sick Children, Toronto, ON, Canada, [2]Neuromuscular and Neurometabolic Clinic, Hamilton Hlth.Sci., Hamilton, ON, Canada, [3]Sherbrooke Univ., Toronto, ON, Canada, [4]Complexe Hosp de la Sagamie, Chicoutimi, QC, Canada, [5]Dept. of Biology and Med. Genetics, Univ. Hosp. Motol, Prague, Czech Republic, [6]Ontario Power Generation Bldg,, 700 Univ. Ave Rm 3-709, Mount Sinai Hosp., Toronto, Ontario,, Toronto, ON, Canada

Disclosure Block: **V. Li:** None.

Huntingtonâ□□s disease (HD), spinocerebellar ataxias (SCAs), C9orf72-ALS/FTD, and myotonic dystrophy type 1 (DM1), are caused by the expansion of tandem repeats in a disease-specific gene. Larger repeat expansions are associated with earlier age-of-onset and increased disease severity. Mounting evidence implicates a spontaneously elevated DNA damage response (DDR) in the absence of exogenous damage contributing to disease in HD, SCAs, and C9orf72-ALS/FTD. Here we determined whether an abnormal DDR is also present in DM1. Myotonic dystrophy type 1 is divided into classical DM1 and congenital (CDM1) subtypes, with CDM1 being the most severe due in part to a larger CTG tract. Interestingly, there are rare classical DM1 cases where a contracted CTG is transmitted to the next generation, showing delayed disease onset and reduced disease severity. We assessed the spontaneous DDR through gamma-H2AX levels in fibroblasts derived from patients with the various forms of DM1. Greater gamma-H2AX levels were evident in classical DM1 and CDM1 relative to unaffected controls, with the highest in CDM1. In contrast, contraction individuals and their children were not significantly different from controls. An elevated DDR was not repeat length dependent as contraction families had similar repeat sizes relative to classical DM1 and CDM1 cells, and some classical DM1 patients had larger expansions than CDM1 patients. Our findings reveal a spontaneous DDR previously observed in other repeat expansion diseases also exists in DM1, and gamma-H2AX levels correlate with disease severity, rather than repeat size.

PrgmNr 3103 - Variant analysis in Gaucher sibling pairs discordant for parkinsonism to identify secondary genetic risk factors for *GBA1*-associated Parkinson disease

View session detail

Author Block: **N. Tayebi**, E. G. Woo, G. Lopez, B. Stubblefield, E. Sidransky; NIH/NHGRI, Bethesda, MD

Disclosure Block: **N. Tayebi:** None.

Parkinson disease (PD), caused by complex interactions between genes and/or environmental factors, is characterized by the cytoplasmic accumulation of misfolded SNCA protein in specific brain regions. A milestone genetic finding was that variants in *GBA1,* the gene encoding lysosomal glucocerebrosidase (GCase), confer an increased risk for PD. Mutations in *GBA1* cause the rare autosomal recessive lysosomal storage disease, Gaucher disease (GD). However, most GD patients and carriers with *GBA1* mutations do not develop parkinsonism, implicating additional risk factors including other genetic variants, aging genes, genetic background, environment, and epigenetics. In this study we sought to identify secondary genetic risk variants by performing exome sequencing on DNA samples from nine sibling pairs with GD discordant for PD, adding additional family members in two informative nuclear families. Six of the sibling pairs had Ashkenazi Jewish ancestry and three European backgrounds. Exome data was annotated with Annovar. Quality variants (MPG Score â¥ 10, Score/Coverage â¥ 5) were filtered to include those with a gnomAD v2.11 exome allele frequency â¤ 0.05 and predicted to be damaging or likely damaging by at least two out of three in silico predictors: SIFT, PolyPhen, and MutationTaster. To identify variants, genes, and pathways serving as potential secondary risk factors for parkinsonism, variants were compared within each family to identify those found only in the GD sibling with PD. Among the nine pairs, we also evaluated variants found in at least one GD/PD individual and no GD individuals. Six variants were shared by the GD/PD only siblings in three families and seventy variants were shared by the GD/PD siblings in two families. Interestingly, these shared variants included compound heterozygous variants in two genes, *PCDHB8*, found in the GD/PD sibling in two families with European background, and *ZNF737*, found in two Ashkenazi Jewish families. Furthermore, in the two extended nuclear families, variant analysis uncovered two additional genes of interest only in the PD sibling. Protein-protein network analysis demonstrated that some of the identified genes were linked to SNCA and GCase. Identified candidate variants and genes will be validated in a larger GD/PD cohort. This data confirms the utility of Gaucher sib-pairs and nuclear family studies in identifying the risk factors in a complex disease. This approach likely incorporates factors such as epigenetics, genetic background, predisposition of gene variants, and even chromosomal shuffling in a family or population, providing advantages over pooled heterogeneous samples from different populations.

PrgmNr 3104 - X-linked recessive mutations in *PDZD4* are likely associated with neurodevelopment delay and autism spectrum disorder

View session detail

Author Block: **M. Sandler**[1], J. Serra-Vinardell[1], J. Stephen[1], N. Balanda[1], P. Houston[1], S. Nampoothiri[2], W. A. Gahl[1,3,4], M. V. Malicdan[1,3,4]; [1]Section of Human Biochemical Genetics, Med. Genetics Branch, NHGRI, NIH, Rockville, MD, [2]Dept. of Pediatrics, Amrita Inst. of Med. Sci. and Res. Ctr., Kerala, India, [3]NIH Undiagnosed Diseases Program, NHGRI, NIH, Rockville, MD, [4]Office of the Clinical Director, NHGRI, NIH, Rockville, MD

Disclosure Block: **M. Sandler:** None.

Neurodevelopmental disorders (NDD), a group of diseases that affect the development of the central nervous system, can occur from due to a multitude of genetic predispositions and environmental factors. Despite the high frequency of NDDs in children, little is understood about the mechanism by which they occur. Genetic abnormalities are considered the strongest risk factors for NDD and have been demonstrated to affect multiple facets of brain development and function. Functional genomics is attempting to determine the roles of various genes that may be involved in NDD. In this study, we describe a pediatric proband who presented with learning disability, autistic features, and developmental delay. Exome sequencing re-analysis identified a hemizygous truncating nonsense variant in *PDZD4* that segregated with disease by Sanger sequencing. We also confirmed decreased mRNA expression of *PDZD4* gene in proband fibroblasts compared to controls. While we know that PDZD4 protein and mRNA are expressed highly in brain tissue, little is known about *PDZD4* function, except that it shares a common PDZ domain with other protein faÂÂmilies that are abundant and essential in signal transduction systems. PDZ-domain-containing-proteins have been linked to processes such as ion-channel signaling, transport, and neuronal development. Essential information on how these domains influence synapse biology, however, remains unknown. For these reasons, we chose an in vitro human model to investigate the phenotype of the neural network and examine communication among neurons. We used CRISPR to investigate the functions of *PDZD4*, knocking down genes in i^3N, an induced pluripotent cell (iPSC) line that has an inducible Neurogenin-2 expression, allowing for a scalable production of glutamatergic neurons. Using this cell model, we successfully knocked down the transcription of *PDZD4* by up to 95% and performed RNAseq on the iPSC-derived neurons. RNA seq analysis when compared to control showcased over 600 differentially expressed genes. Using pathway analyses, we observed that many of these genes are involved in central nervous system processes, including synaptogenesis and axonal guidance signaling, highlighting the importance of *PDZD4* in nervous system development. Using epitope tagging technology, we are investigating the localization of this gene in neurons. We anticipate that this experimental approach will be useful in establishing the functional impact of X-linked variants in *PDZD4* identified in NDD patients. Our work will be useful in elucidating the role of the *PDZD4* gene and may provide a better understanding of NDD.

PrgmNr 3105 - Cross-disorder study of neuropsychiatric CNVs and polygenic scores in a large health system-based population

View session detail

Author Block: **M. T. Oetjens**[1], R. K. Linner[1], J. Gudknecht[1], J. B. Leader[1], Regeneron Genetics Center, D. H. Ledbetter[2], C. L. Martin[2]; [1]Geisinger, Danville, PA, [2]Autism & Dev.al Med. Inst., Geisinger, Lewisburg, PA

Disclosure Block: **M.T. Oetjens:** None.

Genetic investigations of neuropsychiatric disorders (NPD) including autism spectrum disorder (ASD), intellectual disability (ID), and schizophrenia (SCZ), have implicated rare copy number variants (CNVs) as important contributors to disease. There is significant variability in how NPD manifest between individuals (i.e., variable expressivity) in the presence of a rare CNV. The uncertainty in risk across NPD can be problematic for genetic counseling of probands and their family members. While these rare CNVs are of large individual effect size, heritability analyses of ASD and SCZ have demonstrated that common variants explain most of the overall genetic liability. Biases associated with case ascertainment and differences in experimental design across studies complicate a cross-disorder characterization of rare CNVs and polygenic scores (representing common variants) from the scientific literature. In this study, we investigated the cross-disorder risk of 31 NPD-related recurrent CNVs and six polygenic scores of cognitive traits or psychiatric disorders in 122,370 patient-participants from DiscovEHR, a large health system-based population with paired exome, genotype, and electronic health data. CNV calling from exome data using the CLAMMS algorithm identified a CNV in ~0.8% (N = 933) of the cohort. Five NPD were examined in this cross-disorder study: attention-deficit/hyperactivity disorder (ADHD; 3,779 cases), autism spectrum disorder (ASD; 450 cases), bipolar disorder (BPD; 6,678 cases), intellectual disability (ID; 676 cases), and schizophrenia (SCZ; 920 cases). We observed several well-known associations with specific CNVs and NPD, including 22q11.2 deletion and SCZ (odds ratio [OR] = 27.59, 95% CI: 7.43-102.38) and 16p11.2 deletion and ID (OR = 31.82, 95% CI: 14.96-67.68). Additionally, we identified novel associations and important replications of previous findings, including 17q12 deletion and ADHD (OR = 13.71, 95% CI: 5.96-31.52) and 16p11.2 duplication and BPD (OR = 2.56, 95% CI: 1.35-4.87), respectively. The effect sizes of polygenic scores were smaller per standard deviation (max OR = 1.36, 95% CI: 1.26-1.48) and across percentiles relative to the CNVs, yet they jointly explained more of the population variance in ADHD, BD, and SCZ than the rare CNVs combined. However, rare CNVs explained more variance for ASD and ID. In conclusion, results from our cross-disorder study of individual CNVs provide more precise estimates of NPD risk and may aid in genetic counseling of these variants. Furthermore, our study reveals that polygenic scores are not yet able to identify individuals at risk for an NPD equivalent to a CNV of large effect size.

PrgmNr 3106 - Infant with PMM2-CDG(CDG-1a) masquerading as infectious pericarditis

View session detail

Author Block: **X. Newman**[1], Y. Martinez[1], P. Jayakar[2]; [1]Nicklaus Children's Hosp., Miami, FL, [2]Miami Children's Hosp., Miami, FL
Disclosure Block: **X. Newman:** None.
PMM2-CDG(CDG-1a) is a disorder of protein N-glycosylation caused by mutations of the phosphomannomutase-2(PMM2) gene. It is by far the most frequent Congenital Disorders of Glycosylation (CDG). In this report we described an unique case of PMM2-CDG with early cardiac manifestation which was masqueraded as infectious pericarditis. The patient is 5 month old Hispanic male with unremarkable birth history. He initially presented to the ED due to a febrile seizure at 2 months old age, patient had full sepsis work-up and was found to have positive RPP for rhino-enterovirus, UA positive for leukesterase, followed by urine culture positive for E.coli. LP performed, CSF fluid normal and culture negative. Cardiomegaly was noted on CXR, BNP elevated, normal Troponin. On auscultation, muffled heart sounds, no heart murmur, EKG normal. The echocardiogram demonstrated large pericardial effusion requiring pericardiocentesis. He was treated with steroids for pericarditis and antibiotics for UTI. Pericardial effusion recurred after two days, requiring second pericardiocentesis and placement of cardiac catheter in the pericardial space to continue drainage, fluid re-accumulated and patient developed tamponade physiology, which resolved after repositioning of the catheter. Pericarditis was suspected to be secondary to viral infection. Patient developed thrombosis of left common iliac, external iliac and femoral arteries, treated with Lovenox. On follow up as outpatient, he was noted to have poor weight gain, severe hypotonia, facial dysmorphism, inverted nipples and abnormal fat pads, global delay and recurrent pericardial effusion needing admission. Brain MRI revealed prominence fissures in the posterior fossa suggesting Cerebellar atrophy. Ophthalmology exam was normal. The NGS panel showed two Pathogenic variants, c.357C>A (p.Phe119Leu) and c.422G>A (p.Arg141His), in PMM2 gene. These variants are on opposite chromosomes. The patient fits into PMM2-CDG (CDG-1a). In conclusion, the clinical manifestation of PMM2-CDG showed a very broad spectrum and multi-organ involvement, cardiac manifestation like pericardial effusion and cardiomyopathy are more common in infancy. However this presentation can be misdiagnosed for more common causes of pericarditis, like post infectious pericarditis. We therefore highly suggest that CDG disorders should be considered in the differential diagnosis of children with early onset multisystem presentation, involving neurological, gastrointestinal and cardiac systems.

PrgmNr 3107 - A scalable bioinformatic approach to recover variation in camouflaged gene regions in whole genome sequencing data from the UK Biobank

View session detail

Author Block: **H. A. Hejase**[1], B. A. J. Sarver[1], K. Y. He[1], L. De Muynck[2], E. A. Khramtsova[1], Q. S. Li[3], B. Smets[2], A. Parrado[1], I. Royaux[2], S. Li[1], M. Black[1], S. Lovestone[2]; [1]Janssen Res. & Dev., Spring House, PA, [2]Janssen Res. & Dev., Beerse, Belgium, [3]Janssen R&D, LLC, Titusville, NJ
Disclosure Block: **H.A. Hejase:** None.
The human genome contains >12,000 gene regions that share a high degree of sequence similarity. Short-read aligners have difficulty mapping reads to such non-unique regions, resulting in low coverage, that are often referred to as â□□camouflagedâ□□. As standard copy number variation (CNV) approaches cannot be utilized in camouflaged regions, we developed a bioinformatics workflow that utilizes short-read datasets to recover these regions and confidently call CNV-based allotypes. In our methodologic approach, camouflaged groups were identified by flagging all the highly similar regions to any given gene body region. Reads with mapping quality score Complement Receptor Type 1 (*CR1*), a gene known to be associated with Alzheimerâ□□s Disease (AD), in whole genome sequencing (WGS) data from the UK Biobank. We assessed our ability to recover CNV-based allotypes of differing length in *CR1* by applying this workflow to WGS data from a total of 556 UK Biobank participants representing age- and sex-matched cases (AD defined as any hospital ICD-10 code: F00, F00.0, F00.1, F00.2, F00.9, G30, G30.0, G30.1, G30.8, G30.9) and controls (individuals without clinical diagnoses, self-reported history, or family history of mental, behavioral, and neurodevelopmental disorders and diseases of the nervous system). We recovered the *CR1* camouflaged region and grouped the samples into clusters putatively recovering three *CR1* allotypes: S homozygote, F homozygote, and S/F heterozygote. We further explored the robustness of this approach by replicating our findings in an independent, internal short-read WGS dataset generated using BGIâ□□s DNBseq technology on 544 AD samples. We found that the frequencies of the S, F, and S/F allotypes ranged depending on their ancestry between [0.05, 0.12], [0.56, 0.7], and [0.24, 0.31], respectively. These estimates are consistent with those previously reported but warrant further validation (i.e., long-read sequencing). Our scalable approach can be generalized to call camouflaged CNVs in large-scale sequencing datasets, thereby recovering variation underlying disease etiology that was previously precluded from investigation.

PrgmNr 3108 - Accurate assignment of disease liability to genetic variants using only population data

View session detail

Author Block: **K. Raraigh**[1], J. M. Collaco[1], J. Betz[1], M. Atalar Aksit[1], N. Blau[2], J. Brown[3], H. C. Dietz[4], G. L. MacCarrick[5], L. M. Nogee[1], M. B. Sheridan[6], H. J. Vernon[5], T. H. Beaty[7], T. A. Louis[1], G. R. Cutting[8]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Univ. Children's Hosp., Zurich, ZH, Switzerland, [3]The Ohio State Univ., Columbus, OH, [4]Johns Hopkins Univ Sch. of Med., Baltimore, MD, [5]Johns Hopkins Hosp, Baltimore, MD, [6]Johns Hopkins Univ, Baltimore, MD, [7]Johns Hopkins Univ, Sch PubHlth, Baltimore, MD, [8]Johns Hopkins Sch Med, Baltimore, MD

Disclosure Block: **K. Raraigh:** None.

Widespread use of next-generation sequencing for diagnostic and research purposes has highlighted the challenge of determining the clinical significance of genetic variation. The American College of Medical Genetics (ACMG) and the Association of Molecular Pathology (AMP) devised a hierarchical scoring approach to estimate pathogenicity using diverse data elements and predictive tools. Availability of variant frequency in public repositories prompted us to test the accuracy of predicting pathogenicity using population data alone. Under the *a priori* assumption that pathogenic or benign variants represent separate groups, we applied a Bayesian method to assign variants to one of two distributions using their counts in samples from affected and apparently healthy groups. This approach, termed BayPR, enabled assignment of probabilities to all variants, regardless of frequency, predicted functional effect or presence in general population databases. Application of BayPR to 103 *cystic fibrosis transmembrane conductance regulator* missense variants with pathogenicity assigned by functional testing produced an area under the receiver operating curve (AUC) of 0.99, which exceeded ten commonly used algorithms (AUC range: 0.54 to 0.97). Application of BayPR to expertly curated variants in eight genes associated with seven Mendelian conditions (autosomal recessive: cystic fibrosis, phenylketonuria, and interstitial lung disease; autosomal dominant: Marfan syndrome and Loeys-Dietz syndrome; X-linked: adrenoleukodystrophy and Barth syndrome) assigned â□¥80% disease-causing probability to 1,350 of 1,374 (98.3%) pathogenic or likely pathogenic variants, while 22 of 23 (95.7%) benign or likely benign variants had predicted probabilities â□¤20%. Using *only* variant count data in affected and unaffected population samples, BayPR can accurately estimate disease liability of rare and private variants responsible for Mendelian disorders, thereby facilitating interpretation in situations commonly encountered in clinical settings.

This work is posted as a pre-print on bioRxiv as of 19 April 2021:
https://doi.org/10.1101/2021.04.19.440463

PrgmNr 3109 - Alternative polyadenylation quantitative trait loci (APA-QTLs) in chronic obstructive pulmonary disease

View session detail

Author Block: **A. Saferali**, W. Kim, J. Yun, Z. Xu, R. Chase, M. H. Cho, P. Castaldi, C. P. Hersh; Brigham & Women's Hosp., Boston, MA
Disclosure Block: **A. Saferali:** None.
Rationale: Genome-wide association studies (GWAS) have identified numerous genetic variants associated with complex disease. Alternative polyadenylation (APA) plays an important role in posttranscriptional regulation by allowing genes to shorten or extend 3â□□UTRs which may contain cis-regulatory elements including microRNA or RNA-binding protein binding sites. Here, we investigate SNPs that are associated with alternative polyadenylation (APA-QTLs) to identify novel functions for GWAS variants in chronic obstructive pulmonary disease (COPD).**Methods:** RNA sequencing was performed on whole blood from 3743 subjects in the COPDGene Study and integrated with whole genome sequencing data from the NHLBI TOPMed project. DaPars2 was used to identify and quantify usage of APA sites. Associations between all SNPs within 1000 kb of a gene (cis-) and APA levels were tested using tensorQTL, adjusting for gender, library prep batch, principal components of APA quantifications and principal components of genetic ancestry. A total of 11,869,333 SNPs were tested for association with 32,849 APA sites. COPD-associated SNPs were ascertained from a published GWAS (Sakornsakolpat et al., Nat Genet 2019;51:3).**Results:** In COPDGene whole blood, we identified 32,849 APA sites corresponding to 8,661 unique genes present, after filtering out APA sites with low usage. Genes containing APA sites in blood were enriched for Reactome pathways of interferon signaling, mRNA splicing and antigen processing. We found that a total of 8,881 APA sites were associated with at least one SNP with q-valueHSPA4), G protein subunit alpha 12 (*GNA12*), and GM2 ganglioside activator (*GM2A*). **Conclusions:** Many SNPs were associated with alternative polyadenylation in peripheral blood. Ten lead variants associated with COPD were APA-QTLs, suggesting that analysis of APA in whole blood and other relevant tissues can provide novel insights into disease mechanisms. Colocalization analysis and fine mapping will be important next steps to refine the causal variants involved.

PrgmNr 3110 - Analysis of splice site variation using PEX-DETEX indicates an enrichment of poison exons in individuals with developmental brain disorders

View session detail

Author Block: **A. E. HARE-HARRIS**, K. Kelchner, C. Buoy, S. Kerper, R. Tuttle; Bloomsburg Univ., Bloomsburg, PA
Disclosure Block: **A.E. Hare-harris:** None.
Developmental brain disorders (DBD) encompass a range of neurodevelopmental disorders that exhibit shared genetic etiologies and overlapping symptoms. Whole exome sequence studies have identified pathogenic loss of function variants within the coding regions of more than 500 genes in DBD cases. However, the role of intronic variation in DBD has been largely understudied. Splice site variants (SSVs) that disrupt the dinucleotide sequence (GT) at the splice donor site can disrupt the normal pattern of mRNA splicing by retaining an intron in the mature mRNA transcript. In the case that the retained intron contains a stop codon, a truncated protein product is produced, often resulting in nonsense mediated decay (NMD). This NMD-causing intron is called a poison exon (PE). PEs have been identified in targeted analyses of DBD genes, specifically among individuals with Dravet Syndrome who have SSVs in *SCN1A*. However, to date, no large scale analysis of the prevalence of PEs in DBD cases has been conducted. This study aims to develop and use an automated algorithm, Poison EXon DETEction (PEX-DETEX), to identify and assess the prevalence of PEs in known DBD genes in two large genomic datasets. PEX-DETEX is a scalable automated algorithm designed to identify SSVs that cause a PE from VCF datafiles. The PEX-DETEX pipeline utilizes the neural network SpliceAI, to identify SSVs that are predicted to result the retention of an intron in the mature mRNA transcript. The algorithm incorporates the SSV, retained intron sequence, and any alternative splice sites predicted by SpliceAI into the original transcript sequence. PEX-DETEX translates this sequence to identify PEs by detecting premature stop codons within the alternative transcript. PEX-DETEX is currently being used to compare the prevalence of PEs in 517 DBD genes in case and control individuals from ClinVar and gnomAD (v3.1), respectively. Using known PE variants in *SCN1A* as a confirmatory analysis to validate the efficacy of the algorithm, PEX-DETEX was able to identify PEs with strong accuracy. PE predictions were also validated via visual inspection in the UCSC Genome Browser of SSVs in *PTEN, ARID1B,* and *BRCA2*. Overall, analysis of the prevalence of PEs indicates that ClinVar cases are enriched for PEs in DBD genes compared to control individuals. As seen with *SCN1A,* the results of this study may provide new avenues for functional analyses of PE transcripts as potential pharmacological targets. Furthermore, use of PEX-DETEX in genomic studies has broad implications for the assessment of SSVs that are often categorized as uncertain clinical significance and elucidating the role of intronic variation in human diseases.

PrgmNr 3111 - CNV-Disease Curation and Role of Metabotropic Glutamate Receptors Optimized for Neurodevelopmental Disorders with ParseCNV2

View session detail

Author Block: **J. Glessner**[1], M. Khan[1], J. Li[1], X. Chang[2], Y. Liu[3], P. M. Sleiman[2], H. Hakonarson[4]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]CHOP, Philadelphia, PA, [3]The children's Hosp. of philadelphia CHOP, Philadelphia, PA, [4]Children S Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **J. Glessner:** None.
We first discovered significant metabotropic glutamate receptor (mGluR) copy number variation (CNV) associations in ADHD and autism. Whether the singular base, 98% accuracy based on Taqman validation. Comprehensive genomic variant detection and evidence-based assay content and processing optimization is emphasized. The mGluR network constitutes 273 genes from first and second degree protein-protein interactors of the 8 mGluR genes, 4 of which achieved individual statistical significance in our previous studies (Elia et al, 2011). Subsequent analysis found associations to phenotypes including anxiety, depression, mood disorder, Tourette's, anorexia, conduct disorder, oppositional defiant disorder, and schizophrenia of mGluR CNVs compared to control subjects. ParseCNV2 was launched to address CNV burden, known syndromic region querying (CNV GWAS database) and rare recurrent association. CNV detection has remained an area of heavy emphasis for algorithm development; however, both CNV curation and disease association remains in its infancy. ParseCNV2 allows fast and easy curation and association of CNVs in both population and family-based disease settings. Comparable software to ParseCNV2 are benchmarked based on runtime and results. ParseCNV2 starts with a comprehensive quality control pipeline of both samples and individual CNV calls. Beyond deletions and duplications, complex SV (ALU, LINE1, CNV(multiallelic), SVA, INV, INS) parsing is supported. We additionally implemented the DeepCNV algorithm, a deep learning CNV plot validation automatic scoring and the MONTAGE algorithm for Mosaic CNV detection. While many variant call file (VCF) parsers exist, few if any support the variety of VCF presentations and interpretations of CNV genotypes. mGluR genes have been especially pivotal hubs of psychiatric gene association networks. Here we explore the phenotypic and genotypic spectrum of psychiatric disorders in regards to targeted drug therapies. A refined mGluR gene network was introduced to model the most prevalent CNVs in the ADHD population demonstrating the greatest treatment response in neuropsychiatric symptoms to the mGluR modulator fasoracetam. We present a next-generation approach to CNV association by natively supporting the popular VCF specification for sequencing derived variants as well as SNP array PennCNV format. The ParseCNV2 software is available at https://github.com/CAG-CNV/ParseCNV2.

PrgmNr 3112 - Evaluation of splicing in *F9* and *ADAMTS13* Exonic Synonymous Variants

View session detail

Author Block: **K. Jankowska**[1], U. Katneni[1], D. Meyer[1], N. Hamasaki-Katagiri[1], D. D. Holcomb[1], J. Kames[1], C. Kimchi-Sarfaty[2]; [1]FDA, Silver Spring, MD, [2]FDA CBER, Silver Spring, MD
Disclosure Block: **K. Jankowska:** None.
Synonymous single nucleotide variants (sSNVs) are increasingly recognized as contributors to disease phenotype of multiple disease conditions. Splicing dysregulation is considered the primary phenotype inducing mechanism of sSNVs. Additional contributing mechanisms include altered mRNA folding/stability, miRNA binding, translation rate, and co-translational folding. Our models are *F9* and *ADAMTS13* genes: deficiency of FIX, the *F9* product, may cause hemophilia B and deficiency of ADAMTS13 may cause thrombotic thrombocytopenic purpura (TTP). In addition, low ADAMTS 13 plasma levels are predictors of mortality in COVID-19 patients. Previously, we analysed 6 disease-associated and 56 neutral *F9* sSNVs for their effect on splicing using a series of *in silico* splice site prediction tools, splicing regulatory element (SRE) prediction tools, and *in vitro* minigene assays. In this analysis, the use of prediction tools for both splice site and SRE in tandem provided better prediction but those predictions were not always in agreement with the minigene assays. The net effect of variants on splicing dysregulation is context dependent and can be difficult to predict only by *in silico* prediction tools. In this respect, minigene assay could be a straightforward and reliable tool to assess novel variants for potential splicing effects. Currently, we are performing comprehensive *in silico* evaluation of 367 neutral *ADAMTS13* sSNVs. Our preliminary results indicated that some sSNVs change mRNA folding energy/stability, alter mRNA splicing, disturb miRNA binding sites and impact synonymous codon usage. These results suggest that some sSNVs may affect ADAMTS13 function as SNP or in a haplotype and contribute to the large variability in expression and function levels of ADAMTS13 in normal individuals. In conclusion, our results demonstrate that sSNVs are not silent and may contribute to changes in expression and function without affecting the proteinâ□□s amino acid sequence.

PrgmNr 3113 - Exploring the genotypic and phenotypic significance of *KCNJ1* variants in UK Biobank data using REVEAL: Biobank

View session detail

Author Block: **N. Nguyen**[1], S. Sarangi[2], J. L. Brodsky[1], Z. Pitluk[2]; [1]Univ. of Pittsburgh, Pittsburgh, PA, [2]Paradigm4, Waltham, MA
Disclosure Block: **N. Nguyen:** None.
Elevated blood pressure (hypertension) affects one billion people worldwide and is the most common risk factor for cardiovascular disease. Despite its high heritability, the relative contributions of the vast majority of genetic factors to hypertension have yet to be defined. Major regulators of hypertension include the salt-handling transporters in the kidney, among which the renal outer medullary K^+ (ROMK) channel is of particular interest. ROMK, which is encoded by *KCNJ1*, is an integral membrane potassium channel in the kidney. Loss-of-function mutations in ROMK cause a rare salt-wasting disorder, Bartter Syndrome type II. These mutations lead to the premature degradation of ROMK soon after it is synthesized in the cell. Interestingly, heterozygous carriers of these same mutations are protected from hypertension. We therefore hypothesize that polymorphisms in ROMK that result in channel hyperactivity might predispose these individuals to hypertension. To test this hypothesis, we are utilizing REVEAL™: Biobank, a software analytics platform built upon an array-native computational engine, SciDB, to explore the 200K Whole Exome Sequence data from the UK Biobank (Application ID: 51518) . Our goal is to find associations between relevant phenotypes (e.g., blood pressure, urinary sodium and potassium levels) and *KCNJ1* variants using the gene association algorithms, SAIGE and REGENIE. We are also examining correlations between the variants and linkages and relevant blood pressure-lowering medications. Based on these initial analyses, we have begun to select polymorphisms and examine the efficiency of the corresponding proteins to mature and function in yeast and mammalian cells. Preliminary data using the yeast model indicate that two polymorphisms at two different positions in the ROMK sequence that were identified in the UK Biobank and the NIH TOPMed database differentially affect ROMK activity. Our results indicate that relatively subtle amino acid substitutions result in significantly different phenotypes, which may in turn result in a net increase or decrease in blood pressure. In this era of ever-expanding human genomic databases, the consolidation of whole-genome analysis and candidate-based experimental approaches to systematically evaluate the functional impact of human variants is vital. Moreover, as ROMK has emerged as a new target for antihypertensives that lack negative effects on serum potassium, this project may aid in the development of therapeutic strategies to treat heart disease.

PrgmNr 3114 - Genetic control of mRNA splicing is affected by purifying selection and may be linked to incomplete penetrance

View session detail

Author Block: **J. Einson**[1,2], T. Lappalainen[2], D. A. Glinos[2]; [1]Columbia Univ., New York, NY, [2]New York Genome Ctr., New York, NY

Disclosure Block: **J. Einson:** None.

Common variants affecting mRNA splicing are typically identified by splicing quantitative trait locus (sQTL) mapping and have been shown to be enriched for GWAS signals by a similar degree to eQTLs. However, the specific splicing changes induced by these variants have been difficult to characterize, making it more complicated to analyze the effect size and direction of sQTLs.Furthermore, sQTLs may affect the dosage of LoF variants in their target exons. This scenario is a potential driver of incomplete penetrance. To test our model, we first catalogued sQTLs using RNA-seq and WGS data from GTEx v8, using each exonâ□□s percentage spliced in (PSI) metric as a quantitative phenotype. PSI is an interpretable way of assessing an sQTLâ□□s effect size and direction. In total, we identify 5,196 genes with at least one significant exon across at least 1 of 18 GTEx tissues. With this set of sQTLs, it is more common that the derived alleles decrease (n=2,744) rather than increase (n=2,185) the inclusion of their target exons, but have a lower allele frequency distribution compared to sQTLs that increase exon inclusion (K.S. test $p=1.998Ã□10^{-15}$). This suggests purifying selection is acting on sQTL variants based on their regulatory properties.We also performed colocalization analysis between sQTL and GWAS loci across 18 tissues and 114 GWAS studies. We found many examples of sQTL variants colocalizing with GWAS hits, and found some evidence that sQTL effect size, direction and other properties influence the likelihood of a significant colocalization event. Finally, we tested whether sQTLs modifying inclusion of their target exons may modify the penetrance of rare coding variants on the same haplotype. To this end, we analyzed signs of purifying selection by looking for depletion of high penetrance haplotype configurations in a general population. We tested for depletion of high penetrance haplotypes first using all 838 individuals in GTEx v8 with WGS, and then in the larger Trans-Omics for Precision Medicine (TOPMed) project, which includes whole genome sequencing from 63,420 individuals of European descent. The larger sample size allows us to probe penetrance patterns in ultra-rare variants, more common variants, and in genes with multiple rare variants across individuals. Ultimately, we provide insights into the multiple mechanisms how genetic effects on splicing contribute to patterns of genetic variation in human populations and genetic disease risk for common and rare diseases. This technique could improve our interpretation of the risks associated with genetic variation beyond the exome.

PrgmNr 3115 - Identification of discriminative gene-level and protein-level features associated with gain-of-function and loss-of-function mutations

View session detail

Author Block: **C. Bayrak**; Icahn Sch. of Med. at Mount Sinai, New York, NY
Disclosure Block: **C. Bayrak:** None.

Identifying whether a given genetic mutation results in a gene product with increased (gain-of-function; GOF) or diminished (loss-of-function; LOF) activity is an important step toward understanding disease mechanisms as they may result in markedly different clinical phenotypes. Here, we generated the first extensive database of all currently known germline GOF and LOF pathogenic mutations by employing natural language processing (NLP) on the available abstracts in the *Human Gene Mutation Database*. We then investigated various gene- and protein-level features of GOF and LOF mutations by applying machine learning and statistical analyses to identify discriminative features. We found that GOF mutations were enriched in essential genes, autosomal dominant inheritance, protein binding and interaction domains, whereas LOF mutations were enriched in singleton genes, protein-truncating variants, and protein core regions. We developed a user-friendly web-based interface that enables the extraction of selected subsets from the GOF/LOF database by a comprehensive set of annotated features, and downloading up-to-date versions (https://itanlab.shinyapps.io/goflof/). These results could ultimately improve our understanding of how mutations affect gene/protein function thereby guiding future treatment options.

PrgmNr 3116 - Identification of polyadenylation signals relevant to Mendelian disease variant interpretation

View session detail

Author Block: **H. Shiferaw**[1], C. Hong[2], J. J. Johnston[3], L. G. Biesecker[4]; [1]NIH, Gaithersburg, MD, [2]Natl. Inst.S OF HEALTH, BETHESDA, MD, [3]NIH, Rockville, MD, [4]Bethesda, MD
Disclosure Block: **H. Shiferaw:** None.
Polyadenylation is essential in maintaining nascent mRNA stability. Variants in polyadenylation signal (PAS) hexamers can cause reduced polyadenylation at the normal polyA sites and lead to reduced gene expression or expression of transcripts with aberrant 3â□□ UTR sequences. Only 34 PAS hexamer variants in 20 genes have been associated with Mendelian disorders. Of the 20 genes, 12 have at least one PAS hexamer variant as damaging assertion (â□□DMâ□□) in HGMD. We hypothesize that this is under-representative of this class of variants. Here, we aimed to comprehensively identify clinically important PAS hexamers, which are most commonly AATAAA, but other PAS hexamers are included. Clusters of polyadenylated 3â□□ UTRs in ESTs were identified from the PolyA Site 2.0 and examined for dominant polyA site usage activity (defined by >50% of overall EST representation). We filtered for inclusion of polyA sites with PAS hexamers closest to -21 nucleotides relative to the polyadenylation site. We identified 15,212 PAS hexamers for further examination. To understand constraint in the identified PAS hexamers, we compared the number of variants in the AATAAA and ATTAAA PAS hexamers of interest vs. control sequences in 3â□□ UTR by examining variants in gnomAD population. For quality control, we included regions with â□¥20x coverage in >90% of gnomAD samples and removed regions of low complexity and mappability. The AATAAA and ATTAAA PAS hexamers were significantly more constrained than control nucleotides found upstream of our PAS hexamers (p Â® individuals with variants in these hexamers. Twenty-five of 64 genes with PAS hexamer variants showed alternative polyadenylation, which suggests that variants in PAS hexamers can lead to reduced gene function. Additionally, of the 15,212 PAS hexamers, 4,532 were in disease-associated genes, suggesting that PAS hexamer variants may lead to aberrant transcripts in these genes with clinical implications. Further analyses are ongoing to investigate the clinical effects of PAS hexamer variants. These data reinforce that PAS hexamer sequences are critical in polyadenylation and that variants in these sequences are candidates for pathogenic, disease-associated variation.

PrgmNr 3117 - Integrated analysis of co-expression and exome sequencing to prioritize susceptibility genes for cutaneous melanoma

View session detail

Author Block: **S. L. Yepes Torres**, M. A. Tucker, T. Zhang, H. Koka, Y. Xiao, K. Jones, A. Vogt, L. Burdette, W. Luo, B. Zhu, A. Hutchinson, M. Yeager, B. Hicks, N. D. Freedman, K. M. Brown, S. J. Chanock, A. M. Goldstein, X. R. Yang; NIH/NCI, Bethesda, MD

Disclosure Block: **S.L. Yepes Torres:** None.

The application of whole-exome sequencing (WES) has led to the identification of novel high and moderate-risk variants that contribute to melanoma susceptibility. However, confirming disease-causing variants after the initial discovery remains challenging, mainly due to the lack of recurrent variants in multiple families and the presence of genetic heterogeneity. To address this challenge, we applied a gene co-expression network analysis to prioritize candidate genes identified from WES analysis of 34 melanoma-prone families with at least three affected members sequenced per family (n=119 cases), hypothesizing that genes organized on tissue-specific co-expression network may be functionally related. Co-expression network was constructed from GTEx (sun-exposed and non-exposed skin tissue), skin melanoma from the Cancer Genome Atlas (TCGA), and primary melanocyte cultures from 106 newborns using the Weighted Gene Correlation Network Analysis (WGCNA) approach. We performed module-specific enrichment and focused on modules associated with pigmentation processes since they are the best-studied and well-known risk factors for melanoma susceptibility. We found that pigmentation-associated modules across the four expression datasets examined were enriched for well-known melanoma susceptibility genes plus genes with uncharacterized role in melanoma susceptibility. We also used network properties (gene significance, module connectivity, and network representations) to prioritize genes within pigmentation modules as likely susceptibility genes. Integrating information from co-expression network analysis and variant prioritization (allele frequency, pathogenicity, and disease co-segregation), we identified 37 genes (such *DCT, TYR, TYRP1, TPCN2, TRPM1,* and *EPHA5*) as potential melanoma risk genes in our families. The network analysis approach also allowed us to link families with â□□privateâ□□ gene mutations based on gene co-expression patterns and thereby may provide an innovative analysis perspective in gene identification in high-risk families.

PrgmNr 3118 - Loss-of-function tolerance in human disease genes

View session detail

Author Block: **S. Gudmundsson**[1,2,3], M. Singer-Berk[1,3], K. J. Karczewski[1,3], N. Watts[1], E. G. Seaby[1,2,4], S. Baxter[1], Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, D. G. MacArthur[1,5,6], A. Oâ□□Donnell-Luria[1,2,3]; [1]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Div. of Genetics and Genomics, Boston Children's Hosp., Harvard Med. Sch., Boston, MA, [3]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [4]Genomic Informatics Group, Univ. Hosp. Southampton, Southampton, United Kingdom, [5]Ctr. for Population Genomics, Garvan Inst. of Med. Res., and UNSW Sydney, Sydney, Australia, [6]Ctr. for Population Genomics, Murdoch Childrenâ□□s Res. Inst., Melbourne, Australia
Disclosure Block: **S. Gudmundsson:** None.

Databases of human population genetic variation, such as the Genome Aggregation Database (gnomAD), are generally expected to be depleted from variation with severe effects on health. Predicted loss of function (pLoF) variants, including stop gained, essential splice, and frameshift variants, are rare, highly disruptive, and have important implications in disease biology. The gnomAD dataset provides a valuable opportunity to improve our understanding of the effect of pLoF variants and further elucidate reasons for evasion of pLoF.

Previous analysis revealed that at least 28.7% of all high-quality homozygous pLoF variants from 125,748 exomes and 15,708 genomes (1,245 out of 4,336) were unlikely to cause biological LoF. The most common underlying causes for evasion included presence in a weakly expressed exon, rescue by secondary variants disrupting the nonsense trinucleotide code (multi-nucleotide variants) or frame-restoring indels, and splice variants predicted to be rescued by nearby alternative splice sites. The variants still predicted to result in LoF were found in 1,815 genes where homozygous LoF appears to be compatible with life, defined as â□□LoF-tolerantâ□□ (Karczewski et al., 2020). We further explore the disease-relationship of these genes to improve our understanding of LoF evasion.

Of the 1,815 predicted LoF-tolerant genes, 8.7% (n=158) are associated with autosomal recessive or X-linked phenotypes (AR) in OMIM. Reassuringly, 81% (n=128) of the phenotypes are compatible with presence in individuals in gnomAD, confirming that the list largely consists of genes that are LoF-tolerant. However, 32 variants across 30 genes are associated with severe and lethal phenotypes not expected in gnomAD. Deep re-curation identified further possible reasons for evasion of pLoF, mainly deletion of in-frame an exon that does not alter the reading frame, and possible splice site rescues not predicted by splice prediction tools. In summary, guidelines to identify pLoF variants that are unlikely to result in LoF are important in variant interpretation. Investigation of pLoF variants in the gnomAD dataset in recessive disease-genes reveals common mechanisms of escape. Future analysis aims to reveal additional mechanisms of LoF evasion by curation of pLoF variants in dominant disease genes in gnomAD.

PrgmNr 3119 - LUSTR: A powerful and user-friendly tool to call germline and somatic short tandem repeat variants from short read next-generation sequencing data

View session detail

Author Block: **J. Lu**[1], E. Heinzen[1], M. B. Harms[2], C. De Araujo Martins Moreno[2]; [1]The Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Columbia Univ. Irving Med. Ctr., New York, NY
Disclosure Block: **J. Lu:** None.
Short tandem repeats (STRs) are DNA sequences composed of identical or highly similar short repetitive 1-6 base pair units. STRs incomprise > 3% of the human genome and are present in 90% of human genes. STRs mutate in high rates, and may result in either germline or mosaic STR variants. Pathogenic STR expansions are known to associate with several human neurological disorders, including Huntington disease, amyotrophic lateral sclerosis, fragile X syndrome, and beyond. Next-generation sequencing (NGS) technology now allows STR variant calling using purely computational methods. However, challenges still remain. First, perfect information of target STR loci is often required, which can be difficult to infer, especially in the presence of imperfect repeats. Second, STR variant calling is still limited by the structural complexity of the STR and the length of sequenced reads. Third, most existing tools call only germline STR variants. Although pathogenic mosaic STRs have been reported, post-zygotically acquired pathogenic STR variations remain underinvestigated.

Here we introduce LUSTR as a powerful and user-friendly tool to call both germline and somatic STR variants from short read NGS data. LUSTR improves upon existing STR calling software tools in two major ways. First, LUSTR allows flexible input information for target STR loci diminishing the impact of imperfect inputs. This allows for easier user customized target lists, unbiased compilations of known STRs for genome-wide scans, or in any user interested reference sequence. Second, by applying a two-module strategy that first realigns and annotates every single read found to overlap with target STR loci, LUSTR then estimates both size and allele fraction at each STR locus in the outputs to enable subsequent somatic variant analysis. In the simulation tests with reads generated from C9orf72 STR locus incorporating mimic sequencing errors, LUSTR realigned >99.9% of the simulated reads and estimated both size and allele fraction matching expectations. LUSTR also successfully called variants at 13 tested STR loci in short reads sequenced samples from Genome in a Bottle project. Furthermore, by in silico mixing samples, we proved that LUSTR is able to call somatic STR alleles with fraction down to 5%. We propose application of LUSTR will reveal more details and facilitate further investigation into pathogenic somatic STR variants in human neurological disorders.

PrgmNr 3121 - Rare variants affecting multi-omic measurements of gene regulation implicated in EKG traits

View session detail

Author Block: **T. Li**[1], M. Arvanitis[1], B. J. Strober[1], B. Ni[1], V. Wang[1], J. K. Bonnie[1], R. Keener[1], H. J. Lin[2], J. I. Rotter[2], J. A. Brody[3], S. R. Heckbert[3], N. Sotoodehnia[3], D. E. Arking[4], A. Battle[1], NHLBI TOPMed consortium; [1]Johns Hopkins Univ., Baltimore, MD, [2]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [3]Univ. of Washington, Seattle, WA, [4]Johns Hopkins Sch. of Med., Baltimore, MD
Disclosure Block: **T. Li:** None.
Biobank-scale whole genome sequencing has identified molecular pathways relevant to disease that are disrupted by rare coding variants. However, interpretation of non-coding rare variants (RV) is still challenging. RVs have been implicated in complex diseases such as atrial fibrillation (AF). AF variants can be analyzed with electrocardiogram (EKG) traits, but mechanisms of associations remain elusive. Here, we leveraged transcriptomic, proteomic, and methylomic data from Multi-Ethnic Study of Atherosclerosis (MESA) in the TOPMed consortium, to evaluate functional RVs that affect EKG traits through expression of nearby genes in each omic signal. For a subset of 1816 European ancestry individuals, we calculated multi-omic z-scores after regressing out confounders. We observed enrichment of RVs near RNA, protein, and methylation signals with outlier z-scores. We used a hierarchical Bayesian model, Watershed, to prioritize functional RVs and assign posterior probabilities of functional effects for each triplet (gene, individual, RV; 104M analyzed). Compared to WGS annotations, our model significantly improved prediction of multi-omic outlier status for individuals with shared RVs (AUROC = 0.65). The model learned genomic features that predict outlier expression (e.g. TF binding for RNA, frameshift variants for protein). There was good correlation between z-scores of nearby genes and corresponding posteriors for RVs shared with GTEx. Watershed improved risk stratification for diverse polygenic traits over common variants. To study contributions of RVs to EKG traits, we built a framework to evaluate the excess of functional RV burden per gene on 5 EKG traits representing cardiac depolarization and conduction (PR, QRS); repolarization (QT, JT); and rate (RR interval). Multi-omic posteriors prioritized complementary genes associated with extreme EKG values. Extreme QRS intervals were associated with protein outliers of *CFC1*, a risk gene for congenital heart disease, and with mRNA outliers of *STUB1*, a ubiquitin ligase in the heart and striated muscle. A Cox proportional-hazards model showed that extreme values in all EKG traits, except JT, were significantly associated with poorer survival. In a mediation model, 40 genes had outlier protein levels potentially associated with poorer survival through extreme QRS values - for example *ARCN1*, part of the coatomer complex involved in developmental disorders. Our comprehensive survey of RVs and multi-omic outlier signals in a European cohort yields mechanistic insights into EKG traits. These methods are a framework for prioritizing functional RVs in precision medicine studies.

PrgmNr 3122 - Regional missense mutational tolerance in 316,810 exomes

View session detail

Author Block: **K. R. Chao**[1,2], K. E. Samocha[3,2,1], K. J. Karczewski[1,2], gnomAD Consortium, A. H. O'Donnell-Luria[1,2,4], H. L. Rehm[1,2], D. G. MacArthur[5,6,1], G. Tiao[1], M. J. Daly[1,2,7]; [1]Broad Inst. of MIT & Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Wellcome Sanger Inst., Hinxton, United Kingdom, [4]Boston Children's Hosp., Boston, MA, [5]Garvan Inst., Darlinghurst, Australia, [6]Murdoch Children's Res. Inst., Melbourne, Australia, [7]Inst. for Molecular Med. Finland, Helsinki, Finland
Disclosure Block: **K.R. Chao:** None.
Missense variation contributes to the genetic burden in common and rare disease, but missense variants produce dramatically different effects depending on their location and specific amino acid substitutions, complicating their interpretation. Most existing tools that assess missense variant deleteriousness (e.g., SIFT, PolyPhen-2, CADD) focus on the effects of the amino acid substitution and specific site without accounting for the characteristics of the surrounding region. The missense badness, PolyPhen-2, and constraint (MPC) score was developed to incorporate information about both missense depletion within a gene and variant-level metrics to provide a more informative prediction of missense variant deleteriousness.

Here we describe an update to the MPC score and regional missense constraint (RMC) metrics calculated on the gnomAD v2.1 dataset (125,478 exomes). We describe two major refinements to the RMC methods: an improved model of expected missense variation which accounts for DNA methylation and incorporates per-base coverage correction, and increased breakpoint resolution (moving from per-exon to per-base breakpoints). We also test our improved method on the UK Biobank (UKBB) dataset (191,062 exomes) and find similar RMC results between the two datasets. The code for the new RMC pipeline will be released in an open source GitHub repository, and we will upload the updated regions of missense constraint and MPC scores to the gnomAD browser.

The updated gnomAD constraint metrics show 2,671 transcripts with evidence of RMC. Over half of these transcripts (1,426, 53%) overlap the 2700 transcripts found to have evidence of RMC using the ExAC dataset (60,706 exomes). In addition, the updated results include 1,069 new transcripts exhibiting regional missense depletion and remove 1,212 transcripts previously thought to have RMC due to inadequate coverage correction. Furthermore, a larger proportion of the gnomAD results have multiple regions of varying missense constraint: 36% of the previous 2700 transcripts vs 56% of the updated 2671 transcripts have three or more regions of varying missense constraint.

The increased resolution of the RMC results increases the specificity of the MPC score and helps identify variants that fall within previously unappreciated constrained regions. In addition, the updated results aid in clinical interpretation of missense variants, particularly of variants falling within regions of intermediate missense depletion or within regions previously erroneously identified as constrained.

PrgmNr 3123 - Transcription factor analysis across 3,604 GWAS uncovers novel pathway-specific associations

View session detail

Author Block: **C. Breeze**[1], Q. Lan[1], N. Rothman[1], N. Franceschini[2], S. I. Berndt[3]; [1]Natl. Cancer Inst., Bethesda, MD, [2]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [3]Natl. Cancer Inst., Rockville, MD
Disclosure Block: **C. Breeze:** None.
Background: Single nucleotide polymorphisms (SNPs) from genome-wide association study (GWAS) are known to preferentially intersect with active regulatory elements in cell types relevant to disease aetiology. However, aside from a few well-studied SNPs, little is known about transcription factors (TFs) and their binding sites impacted by disease-associated variants. In-depth characterization of TF-specific associations is anticipated to yield new insights into GWAS-mediated disease pathology. However, high false positive rates in TF binding site mapping and a deluge of available epigenomic datasets can hamper researchers working to identify TF-related associations. An organized and efficient analysis strategy is needed to consistently and reliably identify TF-specific associations for GWAS SNPs.
Results: To yield insights into the genetic mechanisms underlying common diseases, we developed FORGE2-TF, an automated web tool for the analysis of TF-specific signal across different GWAS and epigenomic datasets (https://forge2-tf.altiusinstitute.org/). We first used FORGE2 to analyse 3,604 GWAS from the NHGRI-EBI GWAS catalogue using data for DNase I hotspots, 5 histone mark categories and 15 hidden Markov model chromatin states to yield 2,395 tissue-specific enriched GWAS. Of these, 2,363 presented significant trait- or disease-TF enrichments via FORGE2-TF, including > 500 enrichments for immune traits, > 50 enrichments for cardiovascular traits, and > 35 enrichments for neural traits. Importantly, we demonstrate that FORGE2-TF analysis can establish key pathways associated with GWAS-mediated disease aetiology via combined TF and PANTHER pathway analysis adjusting for TF models employed. In addition, a range of newly-associated GO categories were detected, extending our knowledge of tissue-specific regions in GWAS-mediated disease aetiology.
Conclusion: In short, we give a complete list of TF analyses across 3,604 published GWAS, revealing previously unreported associations, pathways and mechanistic insights. Furthermore, we have implemented a straightforward, easy-to-use web tool and browser for TF analysis of further GWAS.

PrgmNr 3124 - Transcriptome-wide association studies and fine-mapping at cell-type resolution

View session detail

Author Block: **H. Shi**[1,2], M. J. Zhang[1,2], K. M. Siewert-Rocks[1,2], S. Gazal[3], A. O. Pisco[4], A. Nathan[5,6,2], S. Raychaudhuri[5,6,2,7], A. L. Price[1,2]; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]USC, Los Angeles, CA, [4]Chan Zuckerberg Biohub, San Francisco, CA, [5]Brigham and Womenâ□□s Hosp., Boston, MA, [6]Harvard Med. Sch., Boston, MA, [7]Univ. of Manchester, Manchester, United Kingdom

Disclosure Block: **H. Shi:** None.

Transcriptome-wide association studies (TWAS) integrating gene expression predictions from cis SNPs with GWAS summary statistics have identified thousands of genes associated to diseases (Gusev et al. 2016 *Nat Genet*, Wainberg et al. 2019 *Nat Genet*). However, TWAS generally use gene expression predictions for bulk tissues, and cannot leverage fine-grained cell-types.

Here, we introduce a powerful approach for performing TWAS at cell-type resolution, leveraging large sample sizes for bulk tissues (GTEx Consortium 2020 *Science*) and high-resolution mouse scRNA-seq data (Tabula Muris Consortium 2020 *Nature*) and the observation that expression of most genes are conserved between human and mouse. We infer cell-type-specific gene expression for each GTEx sample with respect to each Tabula Muris cell type under an empirical Bayes framework, enabling cell-type-specific gene expression prediction and TWAS in each cell type. We also extend gene-level TWAS fine-mapping (Mancuso et al. 2019 *Nat Genet*) that leverages gene expression co-regulations to fine-map causal gene-cell type pairs and to compute posterior mean causal effect sizes in addition to posterior causal probabilities.

We applied cell-type TWAS to GWAS summary statistics for 52 diseases and complex traits (average N=345K), analyzing 343 Tabula Muris cell types from 29 tissues. Across diseases/traits, cell-type TWAS identified a median of 44% more independent gene-disease associations at transcriptome-wide significance (FDRCACNA1C expression in neuronal stem cells with schizophrenia ($P=7Ã□10^{-13}$, vs. $P=3Ã□10^{-8}$ in brain cerebellum in tissue-level TWAS). Although other brain cell types were also significant, TWAS fine-mapping strongly prioritized *CACNA1C* expression in neuronal stem cells over other cell types (posterior causal probability = 1.00, gene expression-disease effect size = -0.02 (s.e. 0.0006)), consistent with known biology. We also identified biological pathways curated in KEGG and Reactome that were associated to diseases using cell-type TWAS. Across diseases/traits, cell-type TWAS identified a median of 30% more pathways (FDRNature).

PrgmNr 3126 - Why do males die? Uncovering the genetic connections causing COVID-19 disease severity

View session detail

Author Block: **M. Hoch**, K. Cardoza, F. Montz, B. Froeschke, K. P. Dobrinski; Univ. of Tampa, Tampa, FL
Disclosure Block: **M. Hoch:** None.
SARS-CoV-2, a randomized killer across global populaces, seems to cause higher death rates within the male population. In an attempt to answer why, this study examines associations with single nucleotide polymorphisms (SNPs) that increase gene expression associated with viral infection and proliferation. ACE2 is believed to be assisted by co-receptors ENPEP and ANPEP with TMPRSS2 serving to cleave the viral protein into two subunits allowing for viral binding to the ACE2 receptor. Due to ACE2 being located on the X chromosome, its expression could potentially be increased or decreased depending on the sex of the COVID-19 patient. The ACE2 gene has SNPs which increase the expression of ACE2. These SNPs were found at frequencies greater than 50% in all male populations analyzed. This could account for the increase in male deaths. Females would undergo X-inactivation for the SNPs and thus have protection from the increased ACE2 expression in all their cells. Population specific SNP patterns were found for ACE2, ENPEP and TMPRSS2 genes which may play a role in increased prevalence of disease among certain populations. Males at higher risk for disease severity would be ones with SNPs for alternate alleles causing increased expression of the ACE2 receptor. Meanwhile, males with the ancestral allele, which is not common globally, would have a decreased expression for the ACE2 receptor which could lead to decreased disease severity. This analysis begins to elucidate the role of personalized medicine in understanding COVID-19 disease susceptibility and severity.

PrgmNr 3127 - CRISPR activation at a trait-associated SNP reveals *trans* effects on gene expression

View session detail

Author Block: **C. J. Cardinale**, X. Chang, H-Q. Qu, X. Wang, C. N. Vaccaro, H. Hakonarson; Children's Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **C.J. Cardinale:** None.
There is currently considerable interest among complex trait geneticists in methods for finding the target genes of trait-associated SNPs discovered in GWAS. We developed a method using CRISPR activation (CRISPRa) in order to find differential gene expression induced by the SNP rs10833518, which we showed is associated with the age-of-onset of type 1 diabetes. We produced capped-and-tailed messenger RNA encoding nuclease-dead Cas9 fused to the histone acetyltransferase domain of p300 (dCas9-p300) by in vitro transcription. This mRNA was electroporated along with synthetic single-guide RNAs (sgRNA) directed to the rs10833518 SNP, or a non-targeting control, in primary human T cells. Total RNA was collected from the cells 3 days post-transfection and subjected to RNA-seq. Although we did not observe cis effects on differential gene expression at the chromosome 11 locus, we did detect trans changes in ODC1 and ARRDC3 messages at approximately 1.5-fold. ODC1 encodes a metabolic enzyme, ornithine decarboxylase. ARRDC3 encodes alpha-arrestin, a signaling mediator that interacts with G protein-coupled receptors. These results show that CRISPRa can reveal modifications to the transcriptome in cases where eQTL or epigenetic analysis of a GWAS locus are unsuccessful.

PrgmNr 3128 - Insights into Molecular Pathways Associated with Juvenile Myositis through Single-Cell Transcriptome Profiling

View session detail

Author Block: **C. Kao**[1], A. Doostparast[1], M. Gonzalez[1], K. Wang[1], H. Hakonarson[1,2]; [1]Ctr. for Applied Genomics, Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Dept. of Pediatrics, Perelman Sch. of Med., Univ of Pennsylvania, Philadelphia, PA

Disclosure Block: **C. Kao:** None.

Inflammatory myopathies (IM)-including dermatomyositis (DM), polymyositis (PM), and inclusion body myositis (IBM)-are a group of rare autoimmune diseases most often characterized by chronic muscle inflammation leading to progressive skeletal muscle weakness and loss of function, where cutaneous involvement may be present (in DM) and systemic problems occur in severe cases. IM presentation is heterogenous, where amyopathic forms (without apparent muscle involvement) can manifest. Juvenile-onset IMs are generally considered distinct but closely related clinical entities from their adult-onset forms. The etiology of IM is unknown, although there is a clear genetic/heritable component. Steroids and other immunomodulatory agents have been transformative, but some cases remain difficult to treat/manage. We are collecting a biobank of samples from families with juvenile IM cases, with the aim of using 'omics' technologies to find potentially informative biomarkers to guide use of existing and possibly new therapies.A number of autoimmune diseases, including DM and PM, show a prominent type I IFN expression 'signature' (~300 genes) which can be tracked in affected tissues and in blood. We previously profiledthe type I IFN signature in blood cells from a group of juvenile IM subjects. Generally, type I IFN activity correlated with disease activity, but a few appear to have persisting disease activity despite a low IFN signature. We performed whole blood RNASeq and single-cell RNASeq on a subset of cases to find potentially informative signals beyond the type I IFN signature. We focus in particular on the mTOR pathway, which has been shown to differentially regulated in monocytes in another autoimmune disorder, Castlemanâ□□s disease, and where targeting mTOR therapeutically (i.e. with rapamycin) is known to be effective in a subset of patients. A number of genes in the mTOR pathway are also differentially expressed in different cell types in juvenile IM, and the potential implications for biomarker development and therapeutic targeting is discussed.

PrgmNr 3129 - Proteomics study on bone biopsy sample

View session detail

Author Block: **F. Yu**[1], S. Reppe[2], K. M. Gautvik[3], Y-H. Hsu[1,4]; [1]Hinda and Marcus Inst. for Aging Res., Hebrew SeniorLife and Harvard Med. Sch., Boston, MA, [2]Dept. of Med. Biochemistry, Oslo Univ. Hosp., Oslo, Norway, [3]Lovisenberg Diakonale Sykehus, Unger-Vetlesen Inst., Oslo, Norway, [4]Broad Inst. of MIT and Harvard, Cambridge, MA

Disclosure Block: **F. Yu:** None.

Accelerated bone loss in older adults increases the risk of osteoporotic fracture, disability, and mortality; and causes a global burden of public health. Proteomics has emerged to be an effective approach to reveal the pathophysiological mechanisms of bone diseases. However, the majority of previous human proteomics studies are on serum/plasma samples. Only a few population-based studies measured proteomics in disease-relevant tissues/cell types. To study proteomics in osteoporosis as well as osteoporotic fracture, we did a comprehensive comparative analysis of proteomes in the pilot samples of a cohort of 200 subjects. Three bone-biopsy samples at three different skeletal sites (caput, collum, and trochanter) and plasma have been collected from each of the three participants with hip replacement surgery or fracture surgery in Oslo University Hospital, Oslo, Norway. The whole proteome of each sample was measured by the high-throughput data-independent acquisition (DIA) liquid chromatography-mass spectrometry (LC-MS/MS). We found that bone samples contain a more diverse proteome compared to plasma samples. The average number of proteins detected are 2069 (caput), 2732(collum), 2721 (trochanter), and 632 (plasma). Each bone site contains a set of unique proteins that are not found in other skeletal sites or plasma samples (percentage of site-specific proteins: 6.7% (caput), 12.2% (collum), 11.8% (trochanter). Bone unique proteins are enriched in some important pathways that are less well-represented by the plasma proteins, including ribosome, Parkinsonâ□□s disease, carbon metabolism, oxidative phosphorylation, etc. Several most abundant bone-classic proteins account for one-third of the total protein mass in bone samples (excluding albumin and hemoglobin), namely, collagen chains, actin, serotransferrin, and vimentin. For the 1775 proteins that are commonly detected in all three bone sites, we performed ANOVA of mixed model and identified 28 proteins that are differentially expressed in specific bone site, including *LRP1*, *NT5E*, and *CS*. In conclusion, by performing the proteomics study in bone samples, we intend to build human bone tissue proteomic atlas that will provide novel insight into the pathophysiological mechanisms of bone health-relevant consequences.

PrgmNr 3130 - Single-cell and spatial transcriptomics of reovirus induced myocarditis

View session detail

Author Block: **M. Mantri**, M. M. Hinchman, S. T. Cross, J. S. L. Parker, I. De Vlaminck; Cornell Univ., Ithaca, NY
Disclosure Block: **M. Mantri:** None.
A significant portion of unexplained deaths in people aged 35 and younger is thought to be due to viral myocarditis, an inflammatory disease of the heart caused by viral infection. The viral and host factors that give rise to injury of the heart following viral infection remain challenging to study, given the heterogeneity of host-pathogen interactions in complex cardiac tissue. Here, we studied the heterogeneous and cell-type specific innate and adaptive immune responses involved in reovirus-induced myocarditis using single-cell and spatial transcriptomics. We infected neonatal C57BL/6 mice with reovirus type-1-lang (T1L) strain (WT) and a mutant strain that does not cause myocarditis (K287T), and applied single-cell and spatial transcriptomic technologies to cells isolated from cardiac, intestinal (primary site of infection) and spleen tissue at three time points post infection. An integrated and time-resolved analysis of single-cell and spatial transcriptomic data revealed stark differences in basal interferon response and innate immune response post infection in cardiac cell types. Pathway enrichment analysis identified a significant upregulation of basal interferon response and innate immune responses four days post infection in endothelial cells. These results demonstrate the importance of endothelial cells lining the cardiac vasculature in initiating host defense to viral infection. Single-cell transcriptomics analysis further revealed a novel Cxcl9-high endothelial cell population involved in the recruitment of T-cells using the Cxcl9-Cxcr3 axis in infected cardiac tissue. Spatial transcriptomics analysis confirmed the co-localization of these signaling cardiac cells and immune cells within the myocarditic patches in the infected tissue. We furthermore found an enrichment of Cxcl9-high endothelial cells as well as Ccl2+ fibroblast cells in WT-infected as compared to K287T mutant infected hearts seven days post infection. This enrichment of signaling by cardiac cells further explained the increased infiltration of immune cells including dendritic cells, T-cells, NK cells, Neutrophils, and eventually, an increased degree of tissue damage in the WT-infected as compared to K287T mutant infected hearts. Together, our results help clarify the role of host-pathogen interactions in reovirus induced myocarditis.

PrgmNr 3132 - ACME: An Affinity-based Cas9 Mediated Enrichment method for targeted nanopore sequencing

View session detail

Author Block: **S. Iyer**[1,2], M. Kramer[1], S. Goodwin[1], W. R. McCombie[1]; [1]Cold Spring Harbor Lab., Cold Spring Harbor, NY, [2]Stony Brook Univ., Stony Brook, NY
Disclosure Block: **S. Iyer:** None.
Structural variations (SV), tend to be recurrent and have been associated with several cancer types. Next-generation sequencing (NGS) is mostly blind to large SVs, lacking sensitivity with false positive rates up to 89% in SV detection. Long-read sequencing generates read lengths of tens of thousands of bases and has helped identify thousands of genomic features pertinent to cancer that were previously missed by NGS. However, the throughput and coverage offered by whole genome long-read sequencing makes it infeasible to conduct large-scale genomic studies. Targeted sequencing significantly improves accuracy and coverage by offering the depth necessary to detect rare alleles in a heterogenous population of cells. Using nanopore Cas9-targeted sequencing (nCATS), a PCR-free enrichment system from Oxford Nanopore Technologies (ONT), we targeted 10 important cancer genes in MCF 10A and SK-BR-3 breast cell lines. However, we observed that the number of reads generated for targets longer than 30 kb were not sufficient to accurately call SVs. To further enhance this approach, we developed an Affinity-based Cas9-Mediated Enrichment (ACME) step, that uses HisTag Dynabeads™ to pulldown non-target reads. With ACME we achieved ~ 75-fold enrichment of the *BRCA2* region and 35-65x coverage of this ~90 kb target on our panel. We observed an increase in enrichment and coverage of other genes on the panel as well, with enrichment as high as 5000-fold for some genes. Across all genes on our panel, we found that ACME helps increase the number of single contiguous reads that span the entire target, which ultimately helps with better alignment and SV detection. ACME could detect all SVs within our target regions that have been previously inferred by PacBio and ONT whole-genome sequencing, but with higher depth. This allows for rare variant detection, making it an effective long-read targeting platform. We are currently testing ACME + native barcoding, to enable targeting a panel of genes across multiple samples, further bringing down per sample sequencing costs. While further optimization is underway, initial testing of this combined approach gave us a mean target coverage of 10-30x. Though unsupported by ONT for the nCATS approach, we have successfully expanded ACMEâ□□s use to the PromethION high throughput device and have observed a 4-fold increase in coverage when compared to a GridION run using the same sample and mass for library prep. With the optimization of the native barcoding approach for ACME, our goal is to perform high throughput long-read targeted sequencing on >15 samples per PromethION flowcell, helping this approach meet production level needs that are currently unmet.

PrgmNr 3134 - Classification of Super populations based on multidimensional scaling analysis of SNP array data

View session detail

Author Block: **H. Qiu**[1], P. Sleiman[1,2], H. Hakonarson[1,2]; [1]Ctr. for Applied Genomics, Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Dept. of Pediatrics, The Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA

Disclosure Block: **H. Qiu:** None.

Population stratification is an important step in genome-wide association studies and researchersâ initial cohort selection often includes a race criteria. In our centerâs in-house web-based information search system - Analyst portal, we implemented classification of super populations (East Asian, South Asian, African, European, American) based on our samplesâ whole genome SNP array data. PLINK 1.9 was used to calculate identity-by-state (IBS) distance matrix and perform multidimensional scaling (MDS). Support Vector Machine (SVM) classifier was used for the downstream super-population classification. Our SNP array datasets were merged with SNP array data from 2,504 individuals in the 1000 Genomes Project (phase 3). We observed that SNP minor allele frequency in the merged datasets had big impact on the final MDS results as visualized by scatter plots. Inclusion of rare SNPs resulted in super population structure not being captured first. Increase in SNP array dataset size increased the computer memory requirement and computation time, and also interfered what the first MDS dimensions actually capture. We set minor allele frequency of 0.1 and chose batch size of SNP array data at 500 samples per batch. We found that the 5 super populations were and accurately separated by the SVM classifier using the first 3 MDS dimensions. The MDS scatter plots remain virtually the same for different data batches genotyped on different SNP arrays. The predicted super population label is used along with self-reported race information from clinical systems, surveys and registries, to make study cohort selection more accurate and flexible.

PrgmNr 3135 - Genomic assessment of single bacteria with Primary Template-directed Amplification for deconvolution of microbial diversity

View session detail

Author Block: R. Carter[1], J. Zawistowski[2], D. Arvapalli[1], **I. Salas-Gonzalez**[1], J. Blackinton[1], J. West[2]; [1]BioSkryb, Raleigh, NC, [2]BioSkryb, Inc., Durham, NC
Disclosure Block: **I. Salas-Gonzalez:** Salary/Employment; BioSkryb.
Taxonomic identification and genomic profiling of individual bacteria within a population provides a level of resolution not possible with metagenomic analysis and has far-reaching applications including host-microbiome interaction, defining species diversity in non-culturable environmental samples, and bioprocessing. A bottleneck to the genomic profiling of single bacteria has been the availability of a robust technology to faithfully amplify the femtogram quantities of DNA of a single bacterium. Here we employ Primary Template-directed Amplification (PTA), a whole-genome amplification technology, to interrogate the genomes of single bacteria isolated by FACS. PTA yields unprecedented genomic coverage uniformity and prevents error propagation by attenuating the size of amplicons, consequently redirecting primers to the primary template and limiting recopying of amplification products. Following lysis of single bacteria, PTA generated 50-200 ng of amplification product with an average amplicon size of approximately 1 kb, for the creation of sequencing libraries by direct adapter ligation of PTA product. We initially verified the ability of our lysis conditions coupled to PTA to yield quality genome assemblies from both gram positive (*B. subtilis*) and gram negative (*E.* coli) bacteria. Contigs assembled from individual *E. coli* cells averaged 100 kb in length while *B. subtilis* contigs averaged 500 kb, highlighting the ability of PTA amplification product to aid the generation of significantly longer contigs relative to multiple displacement amplification (MDA). We next tested the ability of PTA to jointly recover complete high-quality genome assemblies from multiple cells of a mixed culture comprised of *E. coli* and *B. subtilis* cells that were co-lysed and co-amplified. The resulting sub-assemblies of each species from the joint assemblies had cumulative contig lengths that approximated the average genome lengths of the given subspecies, with little evidence for chimeric contigs. In addition to the assembly approach, a phylogenetic marker-based approach also verified the quality of the single-cell genome assemblies. We currently are extending genomic assessment of single bacteria to environmental and clinical microbiome samples. These data demonstrate that PTA can be used to reliably assemble both individual bacterial cells and to jointly assemble small numbers of pooled bacterial cells. This approach enables high-throughput profiling of bacterial populations with the ability to accurately measure granular taxonomic diversity of the population and to characterize the genomic content of individual bacterial species within.

PrgmNr 3136 - Improved first-strand cDNA synthesis kits incorporating qScript Ultra, a novel, engineered, thermostable reverse transcriptase enabling full-length synthesis of cDNA over 20 kb in 10 minutes

View session detail

Author Block: **R. Heller**, D. Schuster; Quantabio, Frederick, MD
Disclosure Block: **R. Heller:** None.

First-strand cDNA synthesis is an indispensable step that underpins many methods for RNA analysis including monitoring changes in gene expression levels, profiling noncoding RNA, RNA pathogen quantification and sequencing of full transcriptome or RNA viruses. Despite the pivotal role of reverse transcriptase (RT), most commercially available RTs have limited speed, processivity, are adversely affected by sample matrix, or lack activity at temperatures sufficient to relax RNA secondary structure. These deficiencies typically result in incomplete and biased cDNA synthesis. Through a program of directed mutagenesis and enzyme variant screening, we have engineered a novel RNase H(-) RT called qScript Ultra that displays superior velocity and processivity with elevated activity over a broad range of temperatures.

The qScript Ultra Flex kit provides qScript Ultra RT in an optimized and stabilized master mix format that delivers efficient full-length cDNA synthesis with maximal yield at temperatures as high as 60Â°C. The enhanced velocity and processivity enables rapid and efficient conversion of long transcripts at elevated temperatures, critical for capturing RNA with high GC-content or regions of secondary structure. These properties are demonstrated by RT-PCR of 20-30 kb fragments using a brief 10-minute RT time, whereas reactions using other engineered RTs produced little to no product. We show linear and consistent cDNA synthesis over a broad range of RNA input quantity and quality, including picogram levels of RNA from FFPE tissues, and samples containing inhibitors derived from extraction chemicals, blood, plant, and animal tissues. qScript Ultra Flex kit provides separate solutions of anchored oligo(dT), randomers, and an enhancer for use with gene-specific primers solutions for flexibility in choice of cDNA priming.

For RT-PCR applications that demand the highest consistency, reproducibility, and unbiased representation of the transcriptome in cDNA product we developed qScript Ultra cDNA SuperMix. Incorporating both anchored oligo(dT) and randomers, this highly stabilized SuperMix contains all required components for cDNA synthesis except RNA template. The improved speed, processivity, and expanded thermal activity profile of qScript Ultra provides linear conversion of RNA to cDNA from 2.5 Âµg to 1 pg of total RNA in a 10-minute reaction at 55-60Â°C. Compared to other cDNA kits, we demonstrate higher conversion efficiencies and reliable RT-qPCR of low abundance transcripts and low copy viral RNA, even in the presence of high levels of carrier RNA.

PrgmNr 3137 - Learning single-cell embeddings for bulk tissue cellular deconvolution

View session detail

Author Block: **S. Sue**[1], M. Huang[1], C. Nagy[1,2], G. Turecki[3], Y. Li[1]; [1]McGill Univ., Montreal, QC, Canada, [2]Douglas Hosp. (McGill Univ.), Montreal, QC, Canada, [3]Mc Gill Univ, Montreal, QC, Canada
Disclosure Block: **S. Sue:** None.

A challenge in studying complex diseases is dissecting its heterogeneity. Part of this challenge is understanding the impact of different conditions on multiple cell types, since it is known that certain cell types are more sensitive to disease. Therefore, the underlying determinants of disease may be partly attributed to changes in cell type composition within the affected tissues. Existing methods such as single-cell RNA-sequencing (RNA-seq), flow cytometry, and immunohistochemistry are generally used as cell counting methods. However, these methods tend to introduce biases during the preparation and processing steps, and tend to be costly when measuring a substantial number of samples. Thus, to address these problems, computational methods have been developed to determine cell proportions using gene expression data. Here, we present our model scETM-Decon, a supervised deep generative model for determining cell type proportions from bulk RNA-seq data by using single-cell RNA-seq data as a reference. The model architecture includes a fully-connected neural network encoder followed by a linear decoder. The linear decoder has a matrix tri-factorization design, comprising of gene, topic, and cell embeddings. The reference single-cell data is used to determine cell type-specific gene signatures, known as our gene embeddings. scETM-Decon is trained on the single-cell data and gene embeddings to learn the encoder network, topic, and cell-specific embeddings. scETM-Decon further accounts for batch-effects in the data as well as gene-specific effects by learning linear intercepts, making the model robust. The trained model is fed the bulk RNA-seq data to determine the resulting bulk cell embeddings. Cell type proportions are then computed by multiplying the single-cell topic embeddings by the bulk cell embeddings as the single-cell topic embeddings tell us how the cell types are represented by each topic while the bulk cell embeddings show how the subjects are represented by each topic. When applied to human pancreatic islet expression data, scETM-Decon outperforms existing state-of-the-art deconvolution methods in accuracy. Preliminary data from deconvolving the human brain prefrontal cortex also show competitive results, suggesting that scETM-Decon will be a reliable and powerful tool for deconvolution. scETM-Decon allows for a more comprehensive understanding of disease, with potential to identify cellular targets that can be used for diagnosis, prevention, and treatment of complex diseases.

PrgmNr 3138 - Long read sequencing reveals extensive structural variation in diverse mouse genomes

View session detail

Author Block: **A. Ferraj**[1,2], P. A. Audano[2], P. Balachandran[2], I. A. Walawalkar[2], A. Czechanski[3], L. G. Reinholdt[3], C. R. Beck[1,2]; [1]Univ. of Connecticut Hlth.Ctr., Farmington, CT, [2]The Jackson Lab. for Genomic Med., Farmington, CT, [3]The Jackson Lab., Bar Harbor, ME
Disclosure Block: **A. Ferraj:** None.
Structural variants (SVs) are genomic rearrangements that are larger than 50 bp in length. SVs consist of insertions, deletions, duplications, inversions, and complex combinations of these events; they collectively have implications in the pathology of cancer, Mendelian diseases, phenotypic variation, and evolution. Whole genome discovery and characterization of SVs has primarily utilized Illumina short-read sequencing, which can lack sensitivity in GC-rich regions and repetitive regions and can falter in the detection of some SVs, including inversions, complex SVs, and insertions. With mobile element polymorphisms, short reads cannot resolve the internal structure and sequence of these events when they are identified. Short reads have been used to study SVs in thousands of human and mouse genomes. In recent years, it has been shown that long-read sequencing excels at identifying many SVs, particularly insertions and duplications; these technologies can also resolve the nucleotide sequence of longer insertions. While the Human Genome Structural Variation Consortium (HGSVC) has released the most complete SV call sets to date, similar long-read efforts have not been applied to mouse strains, which leaves a gap in our understanding of a model organism critical to basic and clinical research. With Pacific Biosciences (PacBio) long-read sequencing, we are annotating and characterizing SVs in 13 widely-utilized inbred laboratory mouse strains. With an ensemble of long read SV calling algorithms, we have identified 495,452 SVs which occur at unique sites across the current mouse reference assembly, cumulatively affecting 10% (300Mb) of the mouse genome. We find that SVs account for over 5x the number of nucleotide changes between strains when compared to single nucleotide variants. Most of novel SVs found (90%) were specific to long-read detection. Furthermore, we have detected strain-specific as well as inter-strain SVs that exhibit high impact consequences, including over 2,300 coding sequence variants, 300 frameshift events, and insertion variants that dramatically affect the transcripts that they occur within. Our data comprehensively identify the extensive SVs present in diverse mouse genomes and strongly suggest that SVs contribute to inbred mouse genomic and phenotypic diversity.

PrgmNr 3139 - MERSCOPE™ reveals the transcriptional organization of the mouse brain

View session detail

Author Block: **G. Emanuel**, J. He, J. L. He; Vizgen, Cambridge, MA
Disclosure Block: **G. Emanuel:** Salary/Employment; Vizgen.
Biological systems are comprised of numerous cell types, intricately organized to form functional tissues and organs. Building molecular atlases to fully understand the structure and function of each cell within the brain is now a key aspect of neuroscience research. Atlas initiatives using single-cell RNA sequencing can characterize cell types based on their RNA expression profiles. However, the tissue organization is lost when cells are dissociated for single-cell sequencing, making it difficult to study how the cellular heterogeneity is contributing to the function of the tissue. Furthermore, accurately characterizing each cell within the brain is challenging due to the low expression of many functionally important genes such as nonsensory G-protein coupled receptors (GPCRs). A true spatial transcriptomics technology with high detection efficiency and single-molecule resolution is required to build accurate and complete molecular atlases. Vizgenâ□□s in situ genomics platform MERSCOPE™ enables the direct profiling of the spatial organization of intact tissue with subcellular resolution. MERSCOPE is built on multiplexed error robust in situ hybridization (MERFISH) technology that uses combinatorial labeling, sequential imaging, and error-robust barcoding to provide the highest detection efficiency and resolution available for spatial genomics. In a single experiment, hundreds of thousands of cells can be spatially profiled with high accuracy and reproducibility. To demonstrate the power of MERSCOPE, we mapped 483 genes across three full mouse brain coronal slices. We constructed a panel of canonical cell type markers and nonsensory GPCRs to spatially profile nonsensory GPCR expression across the brain with cellular context. Nonsensory GPCRs in the brain mediate signaling and may play vital roles behind brain ageing and neurodegenerative disorders but are difficult to analyze. Our experiment successfully detected multiple lowly expressed GPCRs including Oxtr, Tshr, and Insr. The mouse brain receptor map demonstrates MERSCOPE as a leading tool for molecular atlassing, enabling scientists to find greater insights into healthy versus diseased tissue.

PrgmNr 3140 - Normalizing UDI Library Construction for Sensitive Genomic Applications

View session detail

Author Block: **M. Costello**, D. Bays, M. Rahardja, J. Smith, J. Pavlica, J. Leonard, G. Rush, J. Mellor; seqWell, Inc., Beverly, MA
Disclosure Block: **M. Costello:** None.
For a wide range of genomic applications, the ability to sensitively discriminate sequencing reads between different samples is a critical requirement to reduce the effects of noise and index contamination in NGS data. Generally, the method of unique dual-indexing addresses this by adding two sample-specific sequence indices that are unique to each sample. Most available methods that employ UDIs do so by incorporating indexing in the library amplification stage, prior to purification and pooling of libraries for multiplexed sequencing. This represents a high workflow burden because individual libraries have to be carried separately through library construction, amplification, purification, and QC adding significantly to handling time and cost.
Here, we describe a novel approach for UDI library construction that permits pooling of samples immediately after dual-indexed transposase tagging, allowing samples to be purified, amplified, and prepped for sequencing as normalized UDI-tagged multiplexed pools. This new method utilizes a novel decoy-based normalization technology during indexed tn5 transposition that generates uniform quantities of library molecules across a 10-fold range of input DNA from 3 to 30 nanograms, a working range of DNA input amount that is suitable for many common genomic workflows and applications. The 10 bp UDIs were designed to be both greater than two errors and two indels away from any other index in the set and were rigorously screened for performance across multiple sequencing chemistries.
We demonstrate the suitability of this new library construction method on a range of sample types and compare the indexing, normalization, and workflow performance to other commonly used UDI library methods. We anticipate that the method will have wide applicability to NGS workflows that have significant multiplexing requirements that also require the sensitivity and performance benefits of unique dual-indexing.

PrgmNr 3141 - Outcome of 1500 matches through the Matchmaker Exchange for rare disease gene discovery: The 2-year experience of Care4Rare Canada

View session detail

Author Block: **M. Osmond**[1], T. Hartley[1], D. Dyment[1,2], K. D. Kernohan[3], M. Brudno[4,5], O. Buske[6], A. M. Innes[7], K. M. Boycott[1,2]; [1]Children's Hosp. of Eastern Ontario Res. Inst., Ottawa, ON, Canada, [2]Dept. of Genetics, Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada, [3]CHEO, Ottawa, ON, Canada, [4]Hosp Sick Children/Univ Toronto, Toronto, ON, Canada, [5]Techna Inst., Univ. Hlth.Network, Ottawa, ON, Canada, [6]Phenotips, Toronto, ON, Canada, [7]Alberta Children's Hosp, Calgary, AB, Canada
Disclosure Block: **M. Osmond:** None.
Genomic matchmaking, the process of identifying multiple individuals with similar phenotypes and genetic variants of interest, has become an essential practice for discovering novel disease-causing genes in undiagnosed rare disease (RD) patients. The creation of the Matchmaker Exchange (MME), a federated networks of rare disease databases, has made matchmaking more accessible to rare disease stakeholders (like geneticists, genetic counsellors, researchers, and patients), yet little data currently exists on outcomes from using the MME. This study summarizes Care4Rareâ□□s experiences with the MME over the past two years, and highlights lessons that can be applied for future matchmaking endeavors. Patients enrolled as part of the FORGE Canada and Care4Rare Canada research programs over a two year period had their exome sequencing data reanalyzed by a multidisciplinary research team. Compelling variants in genes not previously associated with a human phenotype were flagged for matchmaking, and were submitted through the MME node PhenomeCentral. The genotypes and phenotypes of matched records returned by the MME were reviewed, and unresolved matches were contacted by email for additional comparisons. From July 2018 to July 2020, 194 novel candidate genes were submitted to the MME, resulting in 1514 matches and led to collaborations for 29 genes (15%). About half of these collaborations were due to matching with an already established, unpublished cohort of patients immediately after MME submission. GeneMatcher matches (82% of all matches) almost always required email follow-up due to little information available upfront, while over half of matches from other MME nodes could be ruled out on initial review. Only 56% of these follow-up emails sent for unresolved matches received a response. This study demonstrates that matchmaking through the MME is an effective way to investigate novel candidate genes for undiagnosed RD patients. However, it remains a laborious process, especially when investigating a large number of genes. With most genes in the MME receiving at least one match, more sophisticated approaches to matchmaking will be needed to reduce back-and-forth between users. Such new approaches must therefore emphasize the importance of sharing phenotypic and genotypic data directly with matchmaking platforms.

PrgmNr 3142 - sparQ RNA-Seq Kit: Improved RNA sequencing with 4.5 hr library prep using a novel, integrated ribo-globin depletion technology

View session detail

Author Block: **M. Ait-Ichou**, K. Conley, D. Schuster; Quantabio, Frederick, MD
Disclosure Block: **M. Ait-Ichou:** None.
RNA-seq studies carried out using high-throughput sequencing of cDNA have provided tremendous insight into cellular transcripts studies on a large and more comprehensive scale. However, technical challenges such as laborious and lengthy workflows, affordability, compromised accuracy, read coverage biases, and limited transcript diversity have impeded implementation of the technology in many labs. Here we introduce the sparQ RNA-Seq Kit that integrates RNA fragmentation and depletion of abundant ribosomal and globin transcripts from multiple species (human, mouse, and rat) into a single step, and tube. The proprietary, highly optimized enzymes and streamlined workflow generates high quality directional transcriptome NGS libraries from either intact or degraded RNA samples, with key improvements for low input and FFPE samples, in under 4.5 hours. sparQ RNA-seq libraries were prepared from 1 - 1000 ng of intact total RNA and 10 - 300 ng of FFPE and degraded RNA from blood samples. Library yield, transcript coverage, strand specificity, rRNA and globin depletion rates and gene expression levels were determined and compared with libraries prepared by various RNA-seq kits. The results indicate high correlation between the level of gene detected between high and low input RNA and uniform transcript coverage especially for FFPE and degraded RNA at low RNA input. sparQ RNA-seq libraries also show equivalent or superior performance with regards to library yield and sequencing quality metrics, as well as increased detection of GC-rich transcripts, when compared to libraries prepared by kits from other suppliers. The sparQ RNA-Seq Kit is a RNA-seq library prep kit, with integrated ribo-globin depletion technology, that is affordable, and easy to execute. It enables fastest time to result (4.5 hours), minimal hands-on time (33% less), and less pipetting steps while generating high quality transcriptomic data regardless of sample input, such as cells, FFPE, tissue and whole blood.

PrgmNr 3143 - *TRPC6* gain-of-function in doxorubicin-induced heart failure

View session detail

Author Block: **N. Norton**[1], B. M. Necela[1], X-L. Wang[2], T. Lu[2], H-C. Lee[2]; [1]Mayo Clinic FL, Jacksonville, FL, [2]Mayo Clinic, Rochester, MN
Disclosure Block: **N. Norton:** None.
Introduction: Missense mutations in the non-selective calcium channel, *TRPC6* have previously been shown to be causative of familial focal segmental glomerular sclerosis (FSGS) by gain-of-function. More recently common and rare *TRPC6* variants were associated with doxorubicin-induced cardiomyopathy (CM) and heart failure (HF) in cancer patients, but the mechanism is unknown. A very rare variant, N338S, identified in a single cancer patient with dox-induced HF has not previously been identified in FSGS patients. **Methods:** The mutation of *TRPC6* N338S was introduced using the QuickChange site-directed mutagenesis kit. *TRPC6* wild type (WT) and *TRPC6* N338S mutant were transiently transfected into HEK293 cells using the Effectene transfection reagent kit. After 48-h transfection, whole-cell currents of *TRPC6* WT and N338S mutant were continuously recorded at 30-s intervals from a holding potential of -60 mV to a voltage clamp protocol -100 mV to +100 mV using patch clamp technique. Intracellular Ca^{2+} concentration, $[Ca^{2+}]i$, was measured by Ca^{2+}-sensitive fluorescent dye (fura-2) and intracellular Ca^{2+} imaging system. **Results:** The *TRPC6* N338S mutant showed significantly higher inward and outward current density at voltages of -100 mV and +100 mV respectively, as well as $[Ca^{2+}]i$ in HEK293 cells. Application of 1-oleoyl acetyl-sn-glycerol (OAG, 50 uM) robustly increased the current density and $[Ca^{2+}]i$ of *TRPC6* N338S mutant, compared to those of WT controls. Moreover, a 24-h preincubation with doxorubicin (Dox, 0.5 mM) potentiated the effects of *TRPC6* N338S on $[Ca^{2+}]i$, which were abolished by incubation with 1 uM BI-749327 (a TRPC6 inhibitor). **Conclusions:** Rare missense variants at *TRPC6* may cause doxorubicin-induced heart failure by gain-of-function through increased electrical current and intracellular calcium levels. Inhibitors of TRPC6 are a potential cardioprotective therapy for cancer patients with TRPC6 risk variants should they require doxorubicin.

PrgmNr 3144 - Paradigm Shift from Disease PRS to PGx PRS for Drug Response Prediction using PRS-PGx Methods

View session detail

Author Block: **S. Zhai**, H. Zhang, J. Shen; Biostatistics and Res. Decision Sci., Merck Res. Lab., Merck & Co., Inc., Rahway, NJ
Disclosure Block: **S. Zhai:** None.
Polygenic risk score (PRS), by combining many small prognostic genetic effects, has shown promise in predicting human diseases and complex traits. In efficacy-based pharmacogenomics (PGx) studies, the current practice is to apply such PRS built from disease GWAS directly to PGx data from randomized clinical trials for drug response prediction and patient stratification. However, this approach relies on the assumption that every variant selected for constructing PRS has a constant ratio between its genotype main and genotype-by-treatment interaction effects, which largely may not be true in real PGx data. A violation of such assumption will make disease PRS explain less heritability of drug responses and thus reduce power in predicting them. Here, we propose the paradigm shift from disease PRS to PGx PRS approaches by simultaneously modeling the prognostic and predictive effects and constructing both PRSs for drug response prediction in PGx. We make this paradigm shift possible by developing a series of novel PRS-PGx methods, including PRS-PGx-unadj (unadjusted), PRS-PGx-CT (Clumping + Thresholding), PRS-PGx-L, -GL, -SGL (Lasso-, Group Lasso-, Sparse Group Lasso-based penalized regression), and PRS-PGx-Bayes (Bayesian regression). In the framework of Bayesian regression, we propose a polygenic prediction method that infers posterior prognostic and predictive effect sizes of SNPs simultaneously using PGx genome-wide association summary statistics and an external linkage disequilibrium (LD) reference panel. By introducing global-local continuous shrinkage priors on SNP effect sizes, our proposed PRS-PGx-Bayes method is more robust to varying relationships between the genotype main and genotype-by-treatment interaction effects. Extensive simulation studies show that PRS-PGx methods generally outperform the current disease PRS (PRS-Dis) methods across a wide range of genetic architectures and PRS-PGx-Bayes is superior to all other PRS-PGx methods. We further apply the PRS-PGx methods to the IMPROVE-IT PGx GWAS data by constructing PGx PRSs via 5-fold cross-validation to predict low-density lipoprotein-cholesterol. The drug response prediction results demonstrate the great improvement of PRS-PGx-Bayes in both prediction accuracy and the capability of capturing the treatment-specific predictive effects over alternative methods.

PrgmNr 3145 - Patterns of*CYP2D6*Phenotypes and the impact on precision medicine: New resource from Medical Genetics Summaries

View session detail

Author Block: **A. J. Malheiro**, M. S. Kane, M. A. Hoeppner, D. R. Maglott, B. L. Kattman; NCBI / NLM / NIH, Bethesda, MD
Disclosure Block: **A.J. Malheiro:** None.
The use of genetics in identifying individuals at risk for adverse drug reactions, toxicity or poor response to medication is a growing area of precision medicine and vital to efficient healthcare delivery. Among genes with known or predicted impact on drug responses, *CYP2D6* is a commonly cited biomarker in the US Food and Drug Administration (FDA) drug labels. Genetic variations can alter CYP2D6 enzyme function. These specific alleles and haplotypes can be classified based on the amount of the encoded enzyme activity. Given their allele combinations, an individual is phenotyped as one of the following metabolizer types: ultrarapid, normal, intermediate, poor, or unknown. The frequency of the various CYP2D6 alleles vary globally and show population-specific distribution. Many resources are available to assess individual variant frequencies, but translating these into actionable pharmacogenetic alleles for clinical application is rarely straightforward. Resources from PharmGKB and PharmVar offer data on specific alleles and drug-gene interactions, but this information is of limited use if the individualâ□□s genotype is unknown. Knowledge of the metabolizer phenotype frequencies in specific populations can assist with deciding whether to test an individual. However, searching the biomedical literature on population-specific allele frequencies for each individual is time consuming. Medical Genetics Summaries (MGS) at the National Center for Biotechnology Information (NCBI) now offers allele-specific data for CYP2D6 phenotype frequencies, curated from the published literature, and a centralized list of reference material for additional allele frequency resources. This resource supplements the MGS on over 50 drugs and the genes involved in individual drug responses. The MGS *CYP2D6* reference page provides an overview of the protein function, genotype-to-phenotype translation from pharmacogenomics authoritative sources, links to additional resources, and brief summaries of published literature on phenotype and genotype frequencies. This information will help clinicians identify those individuals at a higher risk of metabolizer alleles that greatly impact how they respond to a drug and what alleles to prioritize in diagnostic testing. When pharmacogenetic testing is warranted, clinicians can select appropriate test options through the linked resources at the NIH Genetic Testing Registry (GTR). This gene-centric resource aims to provide the busy clinician with a single-site resource to improve the adoption of personalized medical care of individuals at risk for adverse or poor reactions from a wide range of medications.

PrgmNr 3146 - Pharmacogenomic polygenic risk score for clopidogrel responsiveness among Caribbean Hispanics

View session detail

Author Block: **j. Duconge**[1], E. Santiago[2], D. Hernandez[1], M. Monero[1], J. Renta[1], P. GonzÃ¡lez[1], H. NÃºÃ±ez[1], K. Melin[1], S. Scott[3], G. RuaÃ±o[4]; [1]Univ. of Puerto Rico Med. Sci. Campus, San Juan, PR, [2]Univ. of Puerto Rico - Med. Sci. Campus, San Juan, PR, [3]Stanford Univ. Med. Ctr., Palo Alto, CA, [4]UConn Hlth., Farmington, CT

Disclosure Block: **J. Duconge:** None.

This multicenter clinical study was aimed at conducting a targeted pharmacogenomic association analysis of residual on-clopidogrel platelet reactivity in 474 Caribbean Hispanic patients. Platelet reactivity was measured using the VerifyNow P2Y12 assay and clopidogrel resistance was defined as P2Y12 reaction units (PRU) â¥208. Genotyping was performed using the whole-genome InfiniumÂ® MEGA BeadChip array. An ancestry-adjusted, weighted polygenic risk score (*w*PGxRS) was developed to account for the effect of multiple variants on PRU and compared between clopidogrel responders and non-responders. The mean PRU across the study cohort was 173.8Â±68.5 and 33.5% of patients were defined as clopidogrel resistant. Multivariate linear regression showed that 19% of PRU variability was attributed to nine independent predictors, with *CYP2C19**2 (rs4244285) accounting for ~7% of observed PRU variation (*p*PON1 rs662, *ABCB1/MDR1* rs2032582, *PEAR1* rs12041331 carrier status and the interaction between African ancestry and rs12041331 carriers also predicted PRUs in participants (*p*â¤0.05). A clear gene-dose effect was seen between PRU and *CYP2C19**2 genotypes, consistent with previous studies in European patient populations, as well as rs12777823. Importantly, a significant positive correlation was detected between the identified *w*PGxRS (five variants) and PRU among the Hispanic patient population (r_p=0.35, *p*wPGxRS discriminated between non-responders and responders (*p*=0.003), indicating that this multigene-based score is a useful predictor of clopidogrel resistance among Caribbean Hispanics. Findings from this work are expected to help close the gap of knowledge about clopidogrel pharmacogenomics and its clinical implementation in this underrepresented population.

PrgmNr 3147 - Results of a Real-world implementation of a pharmacogenetic-empowered comprehensive medication management (CMM) program

View session detail

Author Block: **J. Shaman**, A. P. Peter; Coriell Life Sci., Philadelphia, PA
Disclosure Block: **J. Shaman:** Salary/Employment; Coriell Life Sciences.
Real-world implementation of pharmacogenetic-empowered comprehensive medication management (CMM) can be a useful tool to deal with polypharmacy issues, a common concern in actively aging populations. For over three years, we have been providing a turnkey, population-scale CMM program that combines genetic testing with expert pharmacy review to ensure safe and effective medication use for Medicare eligible members of a state-run retirement system. We tracked pharmacist interactions with patients through their utilization of GeneDose LIVEâ□¢, a clinical decision support (CDS) tool that displayed medication risks including pharmacogenomic test results, provided in silico testing of alternative medications, and recorded and reported the pharmacist notes and actionable information for a patient and their healthcare providers. Specifically, for each patient, pharmacists created a Medication Action Plan (MAP) in the CDS to communicate recommendations to change a drug regimen or specific medication; to monitor drug effectiveness, side effects, or specific laboratory results; and to be aware of potential future medication-related risks due to pharmacogenetic test results with implications for the risk/benefit profile of commonly prescribed medications. In addition, the medication risks identified by the CDS included drug-drug interactions, contraindications, Beers criteria, anticholinergic burden, FDA â□□black-boxâ□□ warnings, and lifestyle factors. The MAPs contained free-text notes and fielded instructions which were analyzed. Evaluation of data from more than 5,000 consented program participants demonstrated that a significant number (72%) of MAPs created by pharmacists were actionable by the memberâ□□s healthcare provider. Of actionable MAPs, 26% contained immediate discontinuation, substitution, or change in medication administration recommendations. Notably, more than 82% of enrollees with MAPs were taking a medication known to be impacted by pharmacogenomics, and of these, about 60% of MAPs contained actionable information specifically related to pharmacogenomic assay results. A propensity score matched difference-in-difference analysis of administrative medical and pharmacy claims data to assess changes in resource utilization following program implementation is ongoing. The results of the impact of this PGx-empowered CMM program on total medical and pharmacy costs will be presented along with results from a full program evaluation.

PrgmNr 3148 - Haplotype-aware inference of human chromosome abnormalities

View session detail

Author Block: **D. Ariad**[1], S. Yan[2], A. Victor[3], F. Barnes[3], C. Zouves[3], M. Viotti[4], R. C. McCoy[2]; [1]Johns Hopkins Univ., Balitmore, MD, [2]Johns Hopkins Univ., Baltimore, MD, [3]Zouves Fertility Ctr., Foster City, CA, [4]Zouves Fndn. for Reproductive Med., Foster City, CA
Disclosure Block: **D. Ariad:** None.
Extra or missing chromosomes---a phenomenon termed aneuploidy---frequently arises during human meiosis and embryonic mitosis and is the leading cause of pregnancy loss, including in the context of in vitro fertilization (IVF). While meiotic aneuploidies affect all cells and are deleterious, mitotic errors generate mosaicism, which may be compatible with healthy live birth. Large-scale abnormalities such as triploidy and haploidy also contribute to adverse pregnancy outcomes, but remain hidden from standard sequencing-based approaches to preimplantation genetic testing for aneuploidies (PGT-A). The ability to reliably distinguish meiotic and mitotic aneuploidies, as well as abnormalities in genome-wide ploidy may thus prove valuable for enhancing IVF outcomes. We developed a statistical method for distinguishing these forms of aneuploidy based on analysis of low-coverage whole-genome sequencing data, which is the current standard in the field. Our approach overcomes the data sparsity by leveraging allele frequencies and linkage disequilibrium (LD) measured in a population reference panel. The method, which we term LD-informed PGT-A (LD-PGTA), retains high accuracy down to coverage as low as 0.05x and at higher coverage can also distinguish between meiosis I and meiosis II errors based on signatures spanning the centromeres.

After evaluating our method via simulation, we applied it to PGT-A data from 8154 IVF embryos. This allows us to refine original diagnoses of trisomies, identifying 448 chromosomes as possessing a haplotype signature of meiotic error that is enriched among chromosomes 16 and 22. In addition, we demonstrate the use of our method for mapping of meiotic crossovers on trisomic chromosomes, which manifest as transitions between distinct haplotype signatures. This will aid future studies investigating associations between aberrant recombination and aneuploidy. We then continued by uncovering genome-wide abnormalities in ploidy (11 haploid embryos and 12 triploid embryos) that were entirely hidden from standard coverage-based analysis. This finding is particularly important given the major contribution of triploidy to cytogenetically abnormal miscarriages.

In summary, our method complements current approaches for PGT-A, while also offering insight into the origins of chromosome abnormalities in human development.

PrgmNr 3149 - Human milk oligosaccharides protect lung health in breastfed infants of the CHILD Study

View session detail

Author Block: **A. Ambalavanan**[1], L. Chang[1], J. Choi[1], Y. Zhang[2], A. Cicero[1], B. Robertson[3], C. Yonemitsu[3], A. Lamri[4], K. Miliku[5], E. Simons[6], S. E. Turvey[7], P. Mandhane[8], A. B. Becker[5], T. Moraes[9], S. S. Anand[10,11], G. Pare[12,13], D. L. Lefebvre[11], S. Malcom[11], P. Subbarao[9], L. Bode[3], A. B. Meghan[5], Q. Duan[14,15]; [1]Dept. BioMed. and Molecular Sci., Queens Univ., Kingston, ON, Canada, [2]Sch. of Computing, Queens Univ., Kingston, ON, Canada, [3]Dept. of Pediatrics and Larsson-Rosenquist Fndn. Mother-Milk-Infant Ctr. of Res. Excellence, San Diego, CA, [4]Dept. of Clinical Epidemiology and Biostatistics, McMaster Univ., Hamilton, ON, Canada, [5]Manitoba Interdisciplinary Lactation Ctr. (MILC), Childrenâ□□s Hosp. Res. Inst. of Manitoba, Winnipeg, MB, Canada, [6]Dept. of Pediatrics and Child Hlth., Univ. of Manitoba, Winnipeg, MB, Canada, [7]BC Children's Hosp. and Univ. of British Columbia, Vancouver, BC, Canada, [8]Dept. of Pediatrics, Univ. of Alberta, Edmonton, AB, Canada, [9]Dept. of Pediatrics, Hosp. for Sick Children and Univ. of Toronto, Toronto, ON, Canada, [10]McMaster Univ, Hamilton, ON, Canada, [11]Dept. of Med., McMaster Univ., Hamilton, ON, Canada, [12]Hamilton, ON, Canada, [13]McMaster Univ., Hamilton, ON, Canada, [14]Queen's Univ, Kingston, ON, Canada, [15]Dept of BioMed. and Molecular Sci., Queens Univ., Kingston, ON, Canada

Disclosure Block: **A. Ambalavanan:** None.

Breastfeeding provides many indisputable health benefits but its impact on lung health and disease remains unclear. Previous studies typically considered breast milk and breastfeeding as a single homogeneous exposure, which fails to acknowledge the complex and dynamic interactions of the mother-breastmilk-infant â□□triadâ□□ as a co-adapted system where variations in each component will influence the others. Breastmilk contains a myriad of bioactive components such as human milk oligosaccharides (HMOs), which vary in concentration among nursing mothers. This study investigates the modulating effects of HMOs on risk of recurrent wheeze and asthma among breastfed infants. A total of 19 HMOs were quantified using high-performance liquid chromatography in the breastmilk samples of 980 lactating mothers, collected 3-4 months postpartum in the CHILD Cohort Study. Genome-wide single nucleotide polymorphisms (SNPs) were genotyped for mothers and their children using the Illumina HumanCoreExome BeadChip and imputations were performed on the Michigan Imputation server using the Haplotype Reference Consortium data. In the mothers, we conducted genome-wide association studies (GWAS) as well as association analyses of maternal exposures (e.g., diet, birth mode, feeding practices) to identify both genetic and non-genetic determinants of HMOs. In the infants, we used gene-environment interaction analyses to investigate how maternal HMOs affect risk of recurrent wheeze and asthma among breastfed infants.We report that HMO concentrations are associated with known loci on chromosome 19 and novel associations on chromosomes 3 and 10 (*P*-8). In addition, we determined that non-genetic factors such as birth mode, prior children, and feeding practices are associated with HMOs (*P*PP = 0.002). In conclusion, our study reports both genetic and non-genetic determinants of HMOs and that specific breastmilk components potentially modulate the risk of respiratory outcomes during early childhood.

PrgmNr 3150 - Impact of maternal nutrition on transcriptome changes over time during fetal liver development

View session detail

Author Block: **K. D. Zimmerman**[1], J. E. Chan[1], P. W. Nathanielsz[2], M. Olivier[3], L. A. Cox[1]; [1]Wake Forest Sch. of Med., Winston-Salem, NC, [2]Univ. of Wyoming, Laramie, WY, [3]Wake Forest Univ. Sch. of Med., Winston-Salem, NC

Disclosure Block: **K.D. Zimmerman:** None.

Maternal nutrition during pregnancy has been well-documented to impact fetal development. Poor maternal nutrition is a major contributor to adverse pregnancy outcomes and development of intrauterine growth restriction (IUGR). However, the molecular mechanisms underlying IUGR are not well-understood, particularly in humans, because they are challenging to study in pregnant women. Here we study a well-established nonhuman primate model of maternal nutrient restriction (MNR) to study the effects of maternal nutrition on the transcriptome of the developing fetus. Fetal liver tissue was collected at caesarean section at 90 days gestation (dG), equivalent to the midpoint of gestation, 120 dG, 140 dG, and 165 dG. At each of the 90 and 165 dG time points, 8 control samples and 8 MNR samples were collected. At each of the 120 dG time points, 6 control samples and 6 MNR samples were collected. Each group had an even number of males and females. Total RNA was isolated from each sample to prepare RNA-Seq libraries, which were sequenced (2x100 base) on the HiSeq 2500 Sequencing System. Data pre-processing, normalization, and differential expression analysis was computed with *limma-voom*. We modeled gene expression as a function of sex (covariate), MNR status, dG (centered), dG squared, and the interactions between dG (or dG squared) and MNR status. When contrasting the linear interaction between time point and MNR status, *KCTD13* but not the other genes you list demonstrated an increase of mRNA expression over time in the MNR group while expression slightly decreased in the control group (FDR p-value = 0.093). *KCTD13* is involved in purine metabolism and overexpression of this gene has been previously associated with microcephaly and may also play a role in oncogenesis. When contrasting the quadratic interaction between time point and MNR status, *MBTD1*, *DPYS*, *HADHB*, and *ST6GAL1* all demonstrated increased expression during the from 90-20 dG and then decreased between 140-165 dG in the control group while the opposite expression pattern was observed in the MNR group (FDR p-value = 0.075). *HADHB* plays a role in mitochondrial beta-oxidation. ST6GAL1 appears to play a role in carbohydrate metabolism. These results highlight some transcriptional changes during development in fetal liver tissue that are potentially associated with decreased growth subsequent to a nutrient poor intrauterine environment, and may impact long-term health outcomes in the offspring.

PrgmNr 3151 - Increased number of *de novo* mutations in craniofacial regulatory elements in trios with orofacial clefts

View session detail

Author Block: **S. W. Curtis**[1], S. Chung[1], M. Bishop[1], K. K. Diaz Perez[1], H. Brand[2], J. C. Murray[3], E. Feingold[4], T. H. Beaty[5], S. M. Weinberg[4], M. L. Marazita[6], M. P. Epstein[1], D. J. Cutler[1], J. Cotney[7], E. J. Leslie[1]; [1]Emory Univ., Atlanta, GA, [2]Boston, MA, [3]U of Iowa, Iowa City, IA, [4]Univ of Pittsburgh, Pittsburgh, PA, [5]Johns Hopkins Univ, Sch PubHlth, Baltimore, MD, [6]Univ Pittsburgh, Pittsburgh, PA, [7]UConn Hlth., Farmington, CT

Disclosure Block: **S.W. Curtis:** None.

Orofacial clefts (OFCs) are the most common craniofacial birth defect, affecting 1 in 700 births. Like other structural birth defects, de novo mutations (DNMs) in coding regions are enriched among OFC probands, especially among genes expressed in craniofacial tissues. A role for noncoding variants is supported by functional follow-up of GWAS signals, but the contribution of noncoding DNMs to the formation of OFCs is not well understood. Therefore, we called and analyzed noncoding DNMs using the whole genome sequencing data in 759 trios in the Gabriella Miller Kids First Research Program with either a cleft lip) or a cleft lip and palate. After calling and filtering DNMs for quality, there were 51,170 DNMs, with an average of 67.4 DNMs per trio. To determine the potential impact, these DNMs were annotated using functional information based on genome-wide epigenetic marks in human embryos 4.5-8 weeks post-fertilization, the critical window for facial development. As a control, DNMs were also called in 329 trios from the 1000 Genome Project (1000GP). While the 1000GP trios had on average more DNMs called (84.3; p

PrgmNr 3153 - Multi omics integration in placenta identifies candidate functional genes for birthweight

View session detail

Author Block: **F. Tekola-Ayele**[1], X. Zeng[1], S. Chatterjee[1], M. Ouidir[1], M. Tesfaye[2], T. Workalemahu[3], R. J. Wapner[4]; [1]Natl. Inst. of Child Hlth.and Human Dev., NIH, Bethesda, MD, [2]NIH, Bethesda, MD, [3]Univ. of Utah, Salt Lake City, UT, [4]Columbia Univ. Irving Med. Ctr., New York, NY
Disclosure Block: **F. Tekola-Ayele:** None.
Abnormal birthweight, in part due to placental dysfunction, has been linked with cardiometabolic and neurological diseases in later life. Genome wide association studies (GWAS) have identified genetic variants associated with birthweight, but the functional mechanisms of the variants remain unclear. The placenta is functionally critical in fetal growth but is still missing from many genomics databases including the Genotype-Tissue Expression portal (GTEx). The goal of this study was to provide functional mechanistic insight into the causal pathway from a genetic variant to birthweight by integrating placental methylation and gene expression with established GWAS loci for birthweight. Placental multi-omics data including genotype, methylation, and RNAseq were obtained from the NICHD Fetal Growth Studies- Singletons (n=301). First, we determined whether birthweight GWAS variants regulate nearby placental gene expression (i.e., expression quantitative trait loci; *cis*-eQTL) and methylation (i.e., methylation quantitative trait loci*; cis*-mQTL). Next, we performed mediation and Mendelian Randomization analyses on variants exhibiting both *cis*-eQTL and *cis*-mQTL to determine causal relationships among the variants, methylation, and gene expression in placenta. Lastly, we applied multi-trait colocalization to investigate whether all three traits (i.e., birthweight, placental methylation, and gene expression) share the same causal variant and to identify candidate genes for functional follow-up. We found that 57.9% (158/273) birthweight GWAS variants were *cis*-mQTL in placenta and 9.5% (26/273) were *cis*-eQTL; 23 variants had co-occurring *cis*-mQTL and *cis*-eQTL effects (false discovery rate WNT3A, *CTDNEP1*, and *RANBP2*) such that the birthweight GWAS variant alters methylation, which in turn causally influences gene expression in placenta. Multi-trait colocalization analysis identified *PLEKHA1*, *FES*, *CTDNEP1*, and *PRMT7* as likely functional effector genes for birthweight (PPA â¥0.75). Our results underscore that placental epigenetic and transcriptomic changes underpin the functional interpretation of several genetic loci associated with birthweight, and potentially inform early origins of diseases in later life.

PrgmNr 3154 - The Effect of Variable Hormonal Replacement Therapies (HRT) on Neurocognition in School-aged Males with 47,XXY

View session detail

Author Block: **M. Brooks**[1], C. A. Samango-Sprouse[2,3,4], M. Hamzik[1], T. Sadeghin[1], A. L. Gropman[5]; [1]The Focus Fndn., Crofton, MD, [2]The Focus Fndn., Davidsonville, MD, [3]George Washington Univ. Sch. of Med. and Hlth.Sci., Washington, DC, [4]Florida Intl. Univ., Dept. of Human and Molecular Genetics, Miami, FL, [5]Children's Natl. Med Ce, Washington, DC
Disclosure Block: **M. Brooks:** None.
Introduction 47,XXY is the most commonly occurring sex chromosome aneuploidy (~1:600 live male births) and is characterized by reduced androgen levels, language-based learning difficulties, speech and motor delay, and executive dysfunction. Recently, hormonal replacement therapy (HRT) has been shown to potentially mitigate some of the endocrine and neurodevelopmental deficiencies associated with a supernumerary X, though optimal administration has remained elusive. The purpose of this study is to delineate the effect of variable HRTs on neurocognitive abilities in a large cohort of school-aged males with 47,XXY. ***Methods*** One-hundred and forty-three prenatally diagnosed school-aged males with 47,XXY (CA: 130 months) were administered HRT based on their pediatric endocrinologistâ□□s assessment: EHT (Early Hormonal Treatment; three intramuscular shots of 25mg testosterone enanthate within the first year of life), HBT (Hormonal Booster Treatment; three shots of 50mg testosterone enanthate from 5 to 8 years of age) and/or TRT (regular low-dose testosterone administration beginning at puberty and continuing thereafter). Each subject received comprehensive neurodevelopmental testing including the Wechsler Intelligence Scale for Children, 4^{th}/5^{th} Edition (WISC-IV/V). Subjects were grouped by HRT status: 43 No-T, 15 EHT, 25 HBT, 22 TRT, 25 EHT-HBT, and 13 EHT-HBT-TRT. A one-way ANOVA and Tukey HSD were utilized to determine differences between treatment groups, if any. ***Results*** A one-way ANOVA revealed there were indeed differences between groups on the WISC-IV/V. Tukey HSD tests revealed significantly higher FSIQ scores among the EHT-HBT-TRT (M=124.00, SD=16.03), EHT-HBT (M=115.77, SD=12.99), and EHT (M=118.27, SD=13.17) groups compared to the No-T group (M=103.68, SD=15.85) (p=0.0002, p=0.02, and p=0.04 respectively). The EHT-HBT group performed significantly better than the No-T group on the Verbal Comprehension (p=0.02) and Processing Speed Indexes (p=0.04). Additionally, the EHT-HBT-TRT group scored significantly higher on the Working Memory Index than the No-T and TRT groups (p=0.00001 and p=0.03, respectively). ***Conclusion*** Consistent with previous findings, males treated with EHT and HBT showed improved working memory capabilities compared to untreated males. Our findings of elevated scores among males treated with any combination of HRTs including EHT suggest this treatment may be essential in optimizing neurodevelopmental outcome in males with 47,XXY. Further research is warranted to delineate the effect of variable HRTs on other neurodevelopmental aspects in this population.

PrgmNr 3155 - The Longitudinal Effect of Early Hormonal Treatment (EHT) in Males with 47,XXY

View session detail

Author Block: **C. A. Samango-Sprouse**[1], M. Brooks[2], S. Song[2], T. Sadeghin[3], A. L. Gropman[4]; [1]George Washington Univ., Washington, DC, [2]The Focus Fndn., Crofton, MD, [3]The Focus Fndn., Davidsonville, MD, [4]Children's Natl. Med Ce, Washington, DC
Disclosure Block: **C.A. Samango-Sprouse:** None.
Introduction 47,XXY is the most frequently occurring sex chromosome aneuploidy, affecting approximately 1 in every 600 males. It is characterized by androgen deficiencies and a variable neurodevelopmental profile including speech and motor delay, language-based learning disabilities, hypotonia, tall stature, and executive dysfunction. Previous research suggests Early Hormonal Treatment (EHT) may improve several aspects of neurodevelopment in infants and young children with 47,XXY. This study investigates whether these positive associations are continued into school-aged males with 47,XXY. ***Methods*** Eighty prenatally diagnosed males with 47,XXY between the ages of 80 and 143 mos were followed through yearly comprehensive neurodevelopmental evaluations. Each was administered assessments based on chronological age including the Wechsler Intelligence Scale for Children 4th/5th Edition (WISC-IV/V), Expressive and Receptive One Word Picture Vocabulary Tests, 4th Edition (EOWPVT-4 and ROWPVT-4), and Leiter International Performance Scale, 3rd Edition (LIPS-III). For statistical analysis, the participants were bifurcated by treatment status: No-T (CA: 101 mos, N=54) and EHT (CA: 108 mos, N=26). Two-tailed t-tests were used to determine statistical differences between groups on each assessment. ***Results*** On the WISC-IV/V, the EHT group scored significantly higher on the Verbal Comprehension Index, Working Memory Index, and FSIQ than the No-T group (p=0.04, p=0.03, and p=0.003 respectively). The EHT group scored an average FSIQ of 116.75 whereas the No-T group scored an average of 102.48. The No-T group (M=109.65, SD=10.85) showed significantly reduced scores on the EOWPVT-4 when compared to the EHT group (M=117.50, SD=6.67) (p=0.03). There was also a significant effect of treatment observed on multiple domains of the LIPS-III including Figure Ground (p=0.001), Classifications & Analogies (p=0.02), Visual Patterns (p=0.05), and Non-Verbal IQ (p=0.05). ***Conclusion*** This is the first study to demonstrate a potential sustained positive impact of early-course androgen treatment on neurodevelopmental capabilities in males with 47,XXY. Consistent with findings in young children, we observed improved expressive language skills in those who received EHT. Overall, our findings elucidate the need for early detection in this population, as neurobiological treatment during critical periods of development may optimize their outcome. With time, more data will become available and studies on the impact of EHT in young adults with 47,XXY may be investigated.

PrgmNr 3156 - Trans-ethnic meta-analysis of genome-wide association studies identifies novel loci for gestational duration

View session detail

Author Block: **S. Chatterjee**[1], J. Wu[1], M. Ouidir[1], K. L. Grantz[1], F. Tekola-Ayele[2]; [1]Epidemiology Branch, Div. of Intramural Population Hlth.Res., Natl. Inst.S OF HEALTH (NIH), BETHESDA, MD, USA, Bethesda, MD, [2]Epidemiology Branch, Div. of Intramural Population Hlth.Res., Natl. Inst.S OF HEALTH (NIH), BETHESDA, MD, USA, BETHESDA, MD
Disclosure Block: **S. Chatterjee:** None.
Shorter gestational duration and preterm birth is the leading cause of morbidity and mortality among neonates. Despite strong evidence that the burden of preterm birth in the United States is disproportionately higher among African Americans, large scale genome-wide association studies (GWAS) are lacking in diverse ancestries. To date, only six loci have been identified in Europeans. To bridge this gap, we performed trans-ethnic GWAS meta-analysis of gestational duration in pregnant women of African (n=637), East Asian (n=238), European (n=622), and Hispanic (n=568) ancestries who participated in the NICHD Fetal Growth Studies. GWAS was conducted in each ancestry on ~ 10 million genotyped and imputed single nucleotide polymorphisms (MAF â¥2% in each population) adjusted for sex, maternal age and five genotype principal components. GWAS summary statistics of each population were combined using trans-ethnic approaches based on random effects accounting for potential genetic effect heterogeneity across diverse populations. The median gestational duration was ~39.4 weeks across all ancestries. Trans-ethnic meta-analysis identified eight genome-wide significant (*P* -8) loci in the maternal genome (*C3, C1QBPP, C4orf36, AFF1, SGCD, CYB5R4, LINC02884, LOC112268135*) associated with gestational duration with consistent effect directions in two or more ancestries. The hits at the complement system-related loci, *C3* and *C1QBPP*, are clinically relevant given prior suggestions that serum *C3* can predict preterm birth. At these loci, each additional minor allele was associated with up to 1-week shorter gestational duration. We also replicated (*P*EBF1 locus identified in a previous study. The loci identified in our study also have annotations implicated in cell differentiation, bone development, cell cycle progression, tumor progression, cardiometabolic traits, and central nervous system development disorders. This study identified novel loci associated with gestational duration via trans-ethnic meta-analysis, including genes encoding the complement system proteins in which maternal serum levels have been associated with preterm birth. Functional follow-up on the loci identified may give clues on clinically relevant molecular intervention targets.

PrgmNr 3157 - Whole mitochondrial genome analysis of a single trophectoderm biopsy - heteroplasmy dynamics and selection in human embryos

View session detail

Author Block: **A. Aggarwal**[1,2], R. Antes[1], S. Madjunkova[1,3], C. Librach[1,2,4,5]; [1]CReATe Fertility Ctr., Toronto, ON, Canada, [2]Dept. of Physiology, Univ. of Toronto, Toronto, ON, Canada, [3]Dept. of Lab. Med. and Pathobiology, Univ. of Toronto, Toronto, ON, Canada, [4]Dept. of Obstetrics and Gynecology, Univ. of Toronto, Toronto, ON, Canada, [5]Inst. of Med. Sci., Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **A. Aggarwal:** None.
Objective: Heteroplasmy (Het) is cellular coexistence of mutated and normal mtDNA copies. Knowledge of effects of mtDNA variants on human embryo development is limited, due to availability of human embryos for research and only recent technology for low DNA input genomic analysis. Whole genome amplification (WGA) is integral to preimplantation genetic testing (PGT); however, it may introduce variant errors and uneven coverage bias, leading to false detection of low-level Mt-Het. We optimized sequencing and analysis of human embryonic mtDNA from a small trophectoderm (TE) biopsy and determined a threshold for detection of heteroplasmy. Our objective was to apply this pipeline to a set of 6 familial cases with maternal-embryonic-fetal mtDNA and study patterns of heteroplasmic variation at the embryonic and fetal stage.**Materials and Methods:** We studied 23 samples from 6 cases. Each case included maternal DNA (blood/granulosa cells), embryonic DNA (WGA DNA from TE biopsy of blastocysts undergoing PGT-A), and fetal DNA of miscarried embryos (product of conception tissue). Illumina DNA Prep with Enrichment using whole MtDNA probe capture by ligation (Twist Biosciences) and Next Seq 550 was used for data generation. Dragen Enrichment was used for analysis. Optimized Het detection threshold of 5% was applied for variant analysis.**Results:** Mt-genome was sequenced at mean depth ~5700x (100% at 50x). 487 variants were identified at 147 sites. 165 Het were detected across 23 samples, and 104 were present at high levels (80%4216T>C in two cases and c.*15928G>A* in one case, both associated with repeated miscarriage on MitoMap. **Conclusions:** Our optimized pipeline allowed detection of low Het variants in human embryonic mtDNA. High variant frequency present in the embryonic mt-genome and is transmitted through a severe germline bottleneck of 27 estimated units. We found a decrease in allele frequency for NS variants from mom to fetus, which is indicative of purifying selection. We are expanding these findings on a set of ~500 embryos. These finding may have key implications for understanding variants in early embryos and their significance in human embryo development.

PrgmNr 3158 - *TESLA*: super-resolution Tumor Edge Structure and Lymphocyte multi-level Annotation from spatial transcriptoms

View session detail

Author Block: **J. Hu**[1], L. Wang[2], M. Li[1]; [1]Univ Pennsylvania, Philadelphia, PA, [2]The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX
Disclosure Block: **J. Hu:** None.
Cancer development is a multistep process involving the accumulation of genetic modifications, which results in the expression of tumor antigens that trigger innate and adaptive antitumor immune responses to eliminate cancer cells. The description of tumor-infiltrating lymphocytes and their metastases is highly related to therapy and survival of patients, which unambiguously demonstrated the importance of the tumor microenvironment (TME) in cancer control. Recent advances in spatially resolved transcriptomics (SRT) technologies have enabled comprehensive characterization of gene expression patterns in the context of the tissue microenvironment. However, experimental methods for SRT can only measure spatial gene expression in discrete spots located apart from each other, leaving large areas unmeasured and limiting their usefulness in studying detailed structure in TME. To overcome this limitation, we present TESLA, an approach for Tumor Edge Structure and Lymphocyte multi-level Annotation at the same resolution as the histology image using SRT data. TESLA integrates morphology information on histology image as well as the gene expression information to annotate different cell types directly on the histology image. Based on the colocalization of some immune cell types, TESLA has the compacity of detecting tertiary lymphoid structures, which is a surrogate marker of a brisk anti-tumor immune response. In addition, TESLA is able to characterize intra-tumor heterogeneity and find genes enriched in the core or edge region of a tumor, representing a promising avenue for understanding the tumor microenvironment.

PrgmNr 3159 - A fast and robust Bayesian nonparametric method for prediction of complex traits using summary statistics

View session detail

Author Block: **G. Zhou**[1], H. Zhao[2]; [1]Yale Univ., New Haven, CT, [2]Yale Univ. Sch. of Publ. Hlth., New Haven, CT

Disclosure Block: **G. Zhou:** None.

Genetic prediction of complex traits has great promise for disease prevention, monitoring, and treatment. The development of accurate risk prediction models is hindered by the wide diversity of genetic architecture across different traits, limited access to individual level data for training and parameter tuning, and the demand for computational resources. To overcome the limitations of the most existing methods that make explicit assumptions on the underlying genetic architecture and need a separate validation data set for parameter tuning, we develop a Summary-statistics based Dirichlet Process Regression method SDPR that does not need to tune parameters. In our implementation, we refine the commonly used likelihood assumption to deal with the discrepancy between summary statistics and external reference panel. Through simulations, we show that SDPR is adaptive to different genetic architectures and robust to heterogeneity of per SNP sample sizes. In real data analysis, we compared the performance of SDPR with 7 recently developed or current state of art methods (PRS-CS, SBayesR, LDpred, P+T, LDpred2, lassosum and DBSLMM) on 6 quantitative (height, BMI, HDL, LDL, total cholesterol and triglycerides) and 6 binary traits (coronary artery disease, breast cancer, IBD, type 2 diabetes, schizophrenia and bipolar). SDPR achieved the best performance for 6 traits (height, BMI, HDL, LDL, total cholesterol, breast cancer), and top tier performance for 4 additional traits (IBD, type 2 diabetes, schizophrenia, bipolar; within 0.003 of AUC difference compared with the top method). Furthermore, SDPR is able to fit the model in 15 minutes when executed in parallel, significantly faster compared with 2-5 hours for PRS-CS, LDpred and LDpred2. Taken together, we believe that SDPR has the potential to be widely used given its competent performance on real traits, easiness to use and excellent computational efficiency. SDPR is freely available at https://github.com/eldronzhou/SDPR.

PrgmNr 3160 - A fresh look at the role of Hardy-Weinberg disequilibrium in association testing

View session detail

Author Block: **L. Zhang**[1], L. J. Strug[2], L. Sun[3]; [1]Univ of Toronto, Toronto, ON, Canada, [2]Hosp for Sick Children, Toronto, ON, Canada, [3]Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **L. Zhang:** None.
Current guidance for Hardy-Weinberg equilibrium (HWE) screening in case-control genome wide association studies (GWAS) is incoherent. First, there is no universal agreement on the HWE p-value threshold as it depends on sample size. Recommended thresholds vary in the literature, ranging from 10E-7 in controls alone (Anderson et al., 2010), 10E-6 in controls and 10E-8 in cases (Marees et al., 2018), and 10E-10 per n=4,000 participants for the UK Biobank study (Bycroft et al., 2018). Second, a truly associated SNP could reasonably be out of HWE in both the case and control groups, even if it is in HWE in the whole population. The degree to which there is Hardy-Weinberg disequilibrium (HWD) due to true association is typically not large but detectable with UK Biobank-type sample size. Consequently, HWE-based quality control may mistakenly screen out truly associated SNPs.
Instead of HWE-based screening and variant removal, we propose a new case-control association test that (a) is robust to genotyping error, (b) leverages HWD attributed to true association to increase power and (c) is easy-to-implement at the genome-wide level. The proposed robust allele-based (RA) joint test expands the recent RA-regression model for conducting robust allelic association analysis (Zhang and Sun, 2021), incorporating the *difference* in HWD between the case and control groups into the traditional association measure. We provide the asymptotic distribution of the proposed test under the null hypothesis of no association. We demonstrate type 1 error control of the RA test at the genome-wide significance level of 5E-8, in the presence of HWD that is attributed to factors unrelated to phenotype-genotype association such as genotype error.
Finally, through a GWAS of meconium ileus in 3,161 individuals (569 cases and 2,592 controls) with cystic fibrosis, we show that the proposed method can (i) robustly analyze SNPs with genotype error, (ii) replicate previous genome-wide significant loci, and (iii) identify novel genome-wide significant loci that were missed by the traditional GWAS approach.

PrgmNr 3161 - A novel regression-based method for X-chromosome-inclusive Hardy-Weinberg equilibrium test

View session detail

Author Block: **L. Sun**[1], L. Zhang[1], Z. Wang[2], A. D. Paterson[3]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Fudan Univ., Shanghai, China, [3]Hosp for Sick Children, Toronto, ON, Canada
Disclosure Block: **L. Sun:** None.
Prior to conducting a genome-wide association study (GWAS), a critical part of the data quality control step is testing the assumption of Hardy-Weinberg equilibrium (HWE), because severe departure from HWE typically indicates genotyping error. The HWE test for an autosomal SNP is straightforward in an independent sample: Contrast the observed genotype counts with the expected under the assumption of HWE and apply the standard Pearsonâ□□s Chisq test for goodness of fit.
How to best perform HWE test for an X-chromosomal SNP, however, is not clear, even using a sample of unrelated individuals. One simple strategy is to use female data only and apply the same Pearsonâ□□s Chisq test. Alternatively, Graffelman and Weir (2016) proposed a 2 d.f. test that includes the deviation of male genotype counts from the expected based on pooled allele frequency estimate using both male and female data.
Instead of the Pearsonâ□□s Chisq-based test, we propose a new regression-based method that (a) analyzes both autosomal and X-chromosomal SNPs, (b) adjusts for covariate effects if needed, (c) analyzes related individuals, (d) includes the existing tests as special cases, and (e) leads to development of new tests. The proposed method builds from the recent robust allele-based (RA) regression model for conducting allelic association test (Zhang and Sun, 2021).
First, we show that a 2 d.f. score test derived from the proposed RA regression includes the test of Graffelman and Weir (2016) as a special case. Second, we show that the 2 d.f. Pearsonâ□□s Chisq test of Graffelman and Weir (2016) can be reformulated as simultaneously testing sex differences in allele frequency, and HWE in female group alone. Thus, we can then develop new HWE tests that do not assume that sex differences in allele frequency are due to genotyping error. These new tests would be more suitable for analyzing, for example, variants that differ in allele frequency due to sex-specific selection. Finally, as the regression approach adjusts for covariate effects, it can analyze samples from multiple populations jointly.
We illustrate the method by application to both phase 3 and high coverage whole genome sequence data from the 1000 genomes project.

PrgmNr 3162 - A sparse high-dimensional generalized varying coefficient model for identifying genetic variants associated with regional methylation

View session detail

Author Block: **K. Zhao**[1], Y. Yang[1], K. Oualkacha[2], C. M. Greenwood[3,1]; [1]McGill Univ., Montreal, QC, Canada, [2]UniversitÃ© de QuÃ©bec Ã MontrÃ©al (UQÃM), Montreal, QC, Canada, [3]Jewish Gen. Hosp., Montreal, QC, Canada

Disclosure Block: **K. Zhao:** None.

DNA methylation is an essential epigenetic modification that regulates gene activity and contributes to tissue differentiation and disease susceptibility. Notably, DNA methylation variation has a vital genetic component. Loci harbouring genetic variants that influence methylation levels are called methylation quantitative trait loci (mQTLs). Identifying mQTLs can provide important insight into the underlying molecular events within multiple human tissues and thus enhance our understanding of the genetic basis of disease development. We have recently proposed a novel varying coefficient model, SOMNiBUS, to analyze regional bisulfite sequencing-derived methylation data, enabling comprehensive and simultaneous estimates of covariate effects that are smoothly varying along genomic positions. However, SOMNiBUS shows important limitations when applying to mQTL analysis because we routinely face hundreds or thousands of candidate SNPs within or near a regulatory region, and SOMNiBUS cannot cope with high-dimensional feature spaces. To address this problem, we propose a high-dimensional varying coefficient model, imposed with a composite penalty term that encourages both sparsity and smoothness for the varying coefficients. We then present an efficient proximal gradient descent algorithm to estimate our model. A comprehensive simulation study has been conducted to evaluate the performance of our approach in terms of estimation, prediction and selection accuracy. We show that our procedure can simultaneously select important mQTLs and estimate their corresponding varying effects across a methylation region of interest with excellent accuracy. An R package called sparseSOMNiBUS that implements our method is freely available on GitHub.

PrgmNr 3163 - AFA Computationally efficient Ancestral Frequency estimation in Admixed populations: the Hispanic Community Health Study/Study of Latinos

View session detail

Author Block: **E. Granot-Hershkovitz**[1,2], S. Browning[3], Q. Sun[4], M. Argos[5], T. Sofer[1,2,6]; [1]Harvard Med. Sch., Boston, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]Univ of Washington, Seattle, WA, [4]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [5]The Univ. of Illinois Chicago, Chicago, IL, [6]Chan Sch. of Publ. Hlth., Boston, MA

Disclosure Block: **E. Granot-Hershkovitz:** None.

Background: Estimation of ancestry-specific allele frequencies in admixed populations is especially relevant for modern-day populations that are becoming increasingly genetically admixed. These allele frequencies are important for prioritizing ancestry-enriched variants for genomic association analyses, inferring demographic histories of populations, interpreting sequence variants, and determining susceptibility to disease. Existing methods for estimating ancestry-specific allele frequencies from an admixed population are time-consuming and limited to a small number of ancestries and, therefore, cannot be applied at scale. **Methods:** We developed a computationally efficient method, Ancestral frequency estimation in Admixed populations (AFA), to estimate the frequencies of bi-allelic variants in admixed populations with an unlimited number of ancestries. AFA uses maximum likelihood estimation by modeling the conditional probability of having an allele given proportions of genetic ancestries, with no need for phased data. It can be applied using either Local ancestry interval proportions encompassing the variant (LAFA) or Global proportions of genetic ancestries (GAFA), which are easier to compute and are more widely available. We evaluated the performance of our proposed method in simulations mimicking admixed populations and implemented the method using the NHLBI BioData Catalyst Powered by SevenBridges, on data from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), an admixed population with three predominant continental ancestries: Amerindian, European, and African. **Results:** Simulations demonstrate that ancestral allele frequency estimation accuracy improves with higher minor allele frequency and larger effective sample sizes of the ancestry of interest. Comparing the HCHS/SOL European and African estimated allele frequencies to their respective gnomAD ancestry frequencies shows high correlations. We defined HCHS/SOL ancestry enriched variants as variants with a minor allele frequency (MAF) of â¥5% in the ancestry of interest and â¤1% in the other two ancestries. Overall, we identified 20,777 Amerindian enriched variants, 135,639 African enriched variants, and 15,063 European enriched variants in chromosome 1. **Conclusion:** Our method can be applied to modern-day admixed populations, enhancing genetic discoveries and personalized genomic health for these understudied populations. We will provide a publicly available code and a dataset of the estimated three ancestral allele frequencies in HCHS/SOL for all available variants with an estimated MAF of â¥5% in at least one of the ancestral populations.

PrgmNr 3164 - Applying Recurrent Weighted Replanting to detect gene-gene interaction in case-parent trios

View session detail

Author Block: **Q. Li**[1], A. Musolf[1], J. E. Bailey-Wilson[2]; [1]Computational and Statistical Genomics Branch, NIH/NHGRI, Baltimore, MD, [2]NIH/NHGRI, Baltimore, MD
Disclosure Block: **Q. Li:** None.
To improve the power to detect causal variants in genomic datasets, a novel procedure called Recurrent Weighted Replanting (RWR) is proposed based on the Random Forest (RF) method. In the past, our group developed trio Random Forest (trioRF) to detect gene-gene interactions in case-parent trio data. TrioRF uses cases and random samples of variant calls from the set of matched pseudo controls, and utilizes a proper classification criterion to detect the difference in variant calls between the pseudo controls and cases. The importance score for each feature is calculated based on permutation tests.
Although trioRF can be scaled up and include a large number of features (SNPs here) in one run to fit a classification tree, the chance of including multiple SNPs in interaction within one run is proportional to the total number of SNPs. As a result, for millions of SNPs, we need to increase the number of trees and improve power. Therefore, we propose to implement RWR for trioRF. We run trioRF multiple times. The initial step is to obtain the importance scores for each SNP (feature). Then we use importance scores from the initial step to select subsets of SNPs to include in subsequent runs and novel weights to adjust the probability that each SNP is available for splitting, a procedure denoted as RWR trioRF. At the final stage, the importance scores from the multiple RWR trioRF runs are calculated. We conducted simulation studies to demonstrate the power of RWR trioRF.

PrgmNr 3165 - Approaches for detection of epistatic interactions of causal variants in genome-wide data: Comparison of Recurrent Weighted Replanting with other machine learning methods

View session detail

Author Block: **J. E. Bailey-Wilson**[1], J. D. Malley[1], A. M. Musolf[2]; [1]NIH/NHGRI, Baltimore, MD, [2]NIH, Baltimore, MD

Disclosure Block: **J.E. Bailey-Wilson:** None.

Much effort is being expended to detect the causes of â□□missing heritabilityâ□□ of complex traits, including effects of rare, moderate to high penetrance risk variants and epistatic interactions. Many machine learning (ML) methods can produce strongly predictive models when the number of potential predictive variables (features) is very large compared to sample size and in the presence of complex interactions between the predictors. However, identifying which of the predictor variables are important in the prediction and therefore biologically important in genetic studies, is not possible for many ML methods. Of those methods that can detect interacting predictors, many do not scale well to analysis of extremely large numbers of features such as millions of genotypes. We have developed r2VIM, based on Random Forests (RF), and have now extended it to Recurrent Weighted Replanting (RWR) to improve power to detect causal variants, with an emphasis on detection of predictors involved in interactions since this is a question of great interest to human geneticists. RWR is a multi-step procedure that runs r2VIM iteratively, using r2VIM importance scores from prior steps to select subsets of features (SNPs here) to include in subsequent runs and novel weights to adjust the probability that each feature is selected as available for splitting (mtry parameter) in any given tree of each RF. At the final stage, r2VIM importance scores are calculated and features that pass a threshold based on importance score variance are selected as important. We simulated case-control data based on a variety of pure epistasis and polygenic plus epistasis models for disease risk, where the interacting features had no main effects since this is the most difficult type of interaction to detect with standard statistical methods such as logistic regression. RWR is powerful and can detect variants involved in epistatic interactions with no marginal effects on the trait (power over 80% across a wide variety of simulated models with 5000 cases, 5000 controls, 100,000 SNPs). We compare these results with observed power from other machine learning methods including RF, r2VIM, Boruta, Vita and TEAM; in general RWR has equal or better power.

PrgmNr 3166 - Characterizing homozygous loss of function variants in 454,782 whole exome sequences from the UK Biobank

View session detail

Author Block: **C. Willis**[1], M. M. Parker[1], A. Deaton[1], L. D. Ward[1], A. Flynn-Carroll[1], R. Hoffing[1], M. Plekan[1], G. Hinkle[1], P. Nioi[2]; [1]Alnylam Pharmaceuticals, Cambridge, MA, [2]Alnylam, Cambridge, MA
Disclosure Block: **C. Willis:** Salary/Employment; Alnylam Pharmaceuticals.
Objective Homozygous predicted loss of function (pLOF) mutations which are expected to inactivate protein-coding genes are valuable when evaluating the therapeutic potential and safety of drug targets. We aimed to characterize high confidence homozygous pLOF variants from 454,782 exomes in the UK Biobank. **Methods** We identified individuals that are either homozygous or heterozygous carriers of pLOF variants (stop-gained, frameshift, or splice donor/acceptor) in 454,782 whole exome sequences from the UK Biobank. Additional variant-level quality control filters were applied to validate both variant calls and functional prediction of LOF including average read depth â¥10, > 98% call rate, and Hardy-Weinberg P-value > $1.0e^{-10}$. Additionally, we only considered autosomes in our analyses. To predict with high certainty whether variants cause loss of function in a gene, only those flagged âhigh confidenceâ by the Loss-Of-Function Transcript Effect Estimator algorithm (LOFTEE) in either the Ensembl MANE (Matched Annotation between NCBI and EBI) or canonical transcript were included. **Results** We identified 3,474 homozygous pLOF variants within 2,260 genes and 618,052 heterozygous pLOF variants in 16,405 genes. The majority of homozygous pLOFs were rare (75% were observed in â¤10 subjects) and nearly half (43%) were only observed in one subject. These unique variants were observed at much higher rates in Southeast Asian (4.3%), African (2.7%), and Chinese (1.9%) ancestry populations than in the European (0.2%) ancestry population. The majority (83%) of identified homozygous pLOF variants were also identified in the gnomAD database with highly correlated allele frequencies (R^2=0.98). More frequent homozygous pLOF variants (n carriers > 1) were more likely to be in a gene that was tolerant to loss of function mutations (pLI = 0). We cross-referenced identified putative knockouts in the UK Biobank with FDA approved drug targets according to the DrugBank database and found 28 genes that overlap. **Conclusions** We identified 2,260 genes with high confidence putative human knockouts in 454,782 exome sequences from the UK Biobank, including known drug targets. Characterization of these knockouts will aid in evaluating the safety and efficacy of genetically defined medicines like RNAi therapeutics.

PrgmNr 3167 - Conditional resampling improves calibration and sensitivity in single-cell CRISPR screen analysis

View session detail

Author Block: **T. Barry**[1], X. Wang[1], J. A. Morris[2], K. Roeder[1], E. Katsevich[3]; [1]Carnegie Mellon Univ., Pittsburgh, PA, [2]New York Genome Ctr., New York, NY, [3]Univ. of Pennsylvania, Philadelphia, PA
Disclosure Block: **T. Barry:** None.
The majority (>90%) of GWAS-implicated loci are thought to lie in *cis*-regulatory elements (CREs). The functional role of most CREs, including the gene or genes through which they exert their effect, is unknown. A central challenge over the coming decade, therefore, is to unravel the *cis*-regulatory landscape of the genome across diseases and cell types.
The most promising technology for mapping candidate CREs to their target genes at genome-wide scale are high multiplicity-of-infection (MOI) single-cell CRISPR screens. Despite their extraordinary promise, high-MOI single-cell screens pose substantial statistical challenges. All published methods designed to analyze data produced by high MOI screens suffer from significant calibration issues, creating excesses of false positive and false negative discoveries. In our work we make two key contributions: we (i) identify core statistical challenges at play in high-MOI single-cell CRISPR screens and (ii) present a novel analysis methodology designed to address them: SCEPTRE (code available at timothy-barry.github.io/sceptre/). SCEPTRE is based on the conditional randomization test, an intuitive and powerful statistical methodology that, like parametric methods, enables simple correction for technical factors such as sequencing depth and batch, and like nonparametric methods, is robust to expression model misspecification. SCEPTRE demonstrated excellent calibration and sensitivity on two recent, large-scale, high MOI single-cell CRISPR screen datasets, revealing hundreds of new regulatory relationships, vigorously validated through a variety orthogonal functional assays.
In a parallel work we combined biobank-scale GWAS data, pooled CRISPR screens, single-cell sequencing, and SCEPTRE to dissect the *cis* and *trans* effects of noncoding variants associated with qualitative blood traits. SCEPTRE confidently mapped 37 noncoding variants to their *cis* target genes â in some cases identifying a causal variant among a set of candidate variants in strong LD â and illuminated several *trans* effects networks. Overall, SCEPTRE is well-positioned to become the method of choice for the wave of noncoding genetic screens likely to be conducted over the next few years.

PrgmNr 3168 - COVID19 symptoms vary by *HLA* haplogroup in a well typed crowd cohort

View session detail

Author Block: **N. M. Pearson**[1], J. Rosenfeld[2], C. Sexton[3], M-L. Endale[3], M. Maasha[3], H. Khiabanian[2], R. Freudenberger[4], J. Feinberg[4], F. Gullo[4], B. Lenes[4]; [1]Root Deep Insight, Inc., Boston, MA, [2]Rutgers Cancer Inst. of New Jersey, New Brunswick, NJ, [3]Root Deep Insight, Boston, MA, [4]Gift of Life, Boca Raton, FL

Disclosure Block: **N.M. Pearson:** None.

Studies of sparsely genotyped people have traced COVID19 susceptibility and severity, in part, to variation in or near human genes encoding cell surface glycoproteins, chemokine and interferon receptors, and other plausibly immune-relevant products. But such common-variant association studies have *not* strongly implicated human leukocyte antigen (*HLA*) genes in SARS-CoV2 response, despite those genes' broad roles in humoral immunity. To better assess how, if at all, diverse and potentially rare *HLA* types may shape COVID19 risks, we invited >200000 largely healthy, well *HLA*-typed bone marrow and stem cell volunteers to answer surveys on exposures, tests, diagnoses, and symptoms during the pandemic. Resulting time-series data, including key covariates on age, sex, locale, background health, and household and kin diagnoses, suggest that variation in or near *HLA* genes may influence COVID19-distinctive symptoms moreso than infection susceptibility per se.

PrgmNr 3169 - Deconvoluting sex-specific effects using GWAS summary statistics and biobank datasets across

View session detail

Author Block: **S. Gao**[1], D. J. Liu[2], C. Khunsriraksakul[3]; [1]Penn State Coll. of Med., Hershey, PA, [2]Penn State Univ., hershey, PA, [3]Hershey, PA
Disclosure Block: **S. Gao:** None.
Many diseases show sex differences, where the disease prevalence and genetic effects of certain causal variants show distinct patterns between sexes. Understanding sex specific genetic architecture is critical for understanding the underlying biology and performing clinical translations. GWAS meta-analyses (GWAMA) aggregate data from multiple studies to enlarge sample size and empower discovery. GWAMA often contain the maximal number of cases and remain the primary source for looking up known associations. In GWAMA, sex is often adjusted as a covariate, but samples of both sexes are often pooled. Very few studies released summary statistics from GWAMA are sex stratified. On the other hand, biobanks, such as UK Biobank, contain a large number of well-genotyped samples, but do not contain enough number of cases for diseases that are not sufficiently common. It is of great interest to jointly analyze these pooled GWAMA summary statistics and biobank datasets to understand sex-specific genetic architecture. To address this, we propose a method MAPBOX (Meta-Analysis Plus Biobanks Or seX-specific analysis) that integrates public GWAMA summary statistics that pool both sexes, with UK Biobank data with sex-stratified results. First, we leverage a subset of variants with clear sex differences to deconvolute the GWAMA summary statistics of pooled sexes and estimate the sex ratio. We then leverage a likelihood framework to jointly model sex-specific effects using both the sex-pooled summary statistics and sex-stratified analysis results from biobanks to estimate sex-specific effects. We conduct extensive simulations and show that MAPBOX is considerably more powerful for identifying sex-specific effects than using sex-stratified biobank alone under a variety of scenarios with different genetic effect distributions, as it combines the sample sizes from GWAMA and biobanks. To apply MAPBOX to real data, we match public GWAMA summary statistics with UK Biobank GWAS results from Neale Lab. For female-specific effects, MAPBOX identifies 382 loci for height, 165 loci for body mass index (BMI) loci and 145 loci for high-density lipoprotein (HDL) loci, while the sex-specific analysis using biobank data only identifies 317 loci for height,144 loci for BMI, 134 loci for HDL. Similar power improvement is also observed for male-specific effects analysis. On average, our method discovers 16% more significant loci in female effect and 19% more in male effect, compared to the method that uses stratified analysis results only. Our method and continued research in this direction will pave the way for more in-depth analysis of sex-specific genetic architectures for complex traits.

PrgmNr 3170 - Developing Trans-ethnic Polygenic Risk Scores Using Empirical Bayes and Super Learning Algorithm

View session detail

Author Block: **H. Zhang**[1], J. Zhan[2], J. Jin[3], J. Zhang[3], T. U. Ahearn[4], Z. Yu[5], J. O. Connell[2], Y. Jiang[2], B. Koelsch[2], 23andMe Research Team, X. Lin[1], M. Garcia-Closas[4], N. Chatterjee[3]; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]23andMe Inc., Sunnyvale, CA, [3]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [4]Natl. Cancer Inst, Rockville, MD, [5]Broad Inst. of MIT and Harvard, Cambridge, MA
Disclosure Block: **H. Zhang:** None.

Polygenic risk scores (PRS) are useful for predicting various phenotypes/outcomes; however, as most PRS have been developed with data generated in European Ancestry (EA) populations, performances of PRS are often poorer in non-EA populations, reflecting their degree of divergence from EA population.

To improve PRS performance in non-EA populations, we propose a novel method, Two-Dimensional Clumping and thresholding with Super Learning and Empirical Bayes (TDLD-SLEB), which takes advantage of both existing large GWAS from EA populations and smaller GWAS from non-EA populations. TDLD-SLEB uses a two-dimensional thresholding method to incorporate SNPs that have either effects in both the larger (e.g., EA populations) and the smaller (e.g., non-EA populations) target population or specific effects in the smaller population. It estimates effect sizes for SNPs in the target population using an Empirical Bayes method that borrows GWAS information across populations. Finally, it incorporates a super learning algorithm to combine series of PRS generated by various SNP selection thresholds for the target population.

Our simulation analyses mimicked real LD patterns using haplotype data of 1000 Genome Phase 3 for five ancestries. We considered various genetic architectures including different levels of negative selection and genetic correlation across ancestries. We found PRSs generated by TDLD-SLEB had significantly improved predictionary accuracy for non-EA populations in independent validation datasets, compared to single ethnic PRS, EUR derived PRS, or a weighted PRS that combines EUR and single ethnic derived PRS with weights selected to optimize prediction in the target population.

Using 23andMe data, we developed and validated population specific PRS for seven complex traits using GWAS data from Europeans (average N=2,442K), African American (average N=113K), Latino (average N=411K), East Asians (average N=94K), South Asians (average N=25K). We found TDLD-SLEB often led to large improvements in the performance of PRS compared to alternative methods for predicting traits in the African American population (average R^2 increased +277% compared to the weighted PRS method). For other ethnic groups, TDLD-SLEB also led to sometimes notable improvements in the performance of PRS, such as for the cardiovascular disease in the Latino population (AUC = 0.61 for TDLD-SLEB vs. AUC= 0.58 for the weighed PRS method).

In conclusion, we developed a computationally scalable and statistically efficient method for generating predictive PRS in non-European populations using GWAS datasets across diverse populations.

PrgmNr 3171 - Ethnicity-specific high-risk gene variant profiling unmasks diabetes associated genes

View session detail

Author Block: **J. Zhang**; NIH, Bethesda, MD
Disclosure Block: **J. Zhang:** None.
The underlying genetic basis for the significant health disparities among ethnic subpopulations and how such inter-ethnic genetic variance may impact the discovery of disease-associated genes warrant a deeper understanding. Through the whole-genome analysis of the total number of high-risk variants (hrV) of each gene in populations representing 185934 subjects, we identified sets of genes with hrV frequencies (hrVF) either unique to each, or conserved in all, ethnicities and used this to develop a quantitative gene-based high-risk variant index (hrVI) of 20428 genes. Gene-to-gene comparisons of ethnicity-specific hrVFs and hrVIs between the case (20,781 subjects) and control (24440 subjects) populations in the type 2 diabetes mellitus (T2DM) national repository identified 57 T2DM-associated genes, 40 of which were discoverable only by ethnicity-specfic analysis. These results demonstrate the utility of hrV analysis and illustrate how gene-based ethnicity-specific analysis of genetic variations can significantly facilitate the identification of genes associated with polygenic diseases

PrgmNr 3172 - Exact and flexible, conditional and joint analysis of association summary statistics using linkage disequilibrium

View session detail

Author Block: **L. G. Sloofman**[1], J. Boocock[2], J. Johnson[3], H. Young[4], W. Wang[3], R. Singer[3], L. Dobbyn[5], T-H. Nguyen[6], E. Stahl[5], L. M. Huckins[4]; [1]Seaver Autism Ctr., New York, NY, [2]Dept. of Human Genetics, Univ. of California, Los Angeles, CA, [3]Pamela Sklar Div. of Psychiatric Genomics, Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Regeneron Genetics Ctr., Tarrytown, NY, [6]Virginia Commonwealth Univ., Richmond, VA

Disclosure Block: **L.G. Sloofman:** None.

Genome-wide association studies (GWAS) have identified thousands of genomic loci associated with hundreds of complex traits and diseases. In order to infer biology from these associations, researchers now seek to identify causal variants within these long lists of significantly associated loci; simply assuming that the lead variant is causal risks overlooking interesting secondary variants, which may still have substantial biological impact. Methods that focus on identifying independent signals within complex loci will be important to elucidating these potentially interesting secondary associations.

Standard conditional analyses to identify these secondary and tertiary associations largely require access to individual-level genotype data. However, the advent of large-scale GWAS from vulnerable or protected populations (for example, 23&Me or MVP analyses), GDPR and other stringent data protection measures, and incomplete sharing of raw data from multi-omic analyses increasingly hamper access to raw data, and standard conditional analyses to refine signals and identify causal variants within genomic loci may not be widely practicable. To address this issue, powerful software exists to perform conditional analyses using summary statistics; most prominently, GCTA-CoJo performs conditional analysis using summary-level statistics and LD estimated from a reference sample. Here, we introduce CoCo, which performs conditional and joint analysis of summary statistics. CoCo offers two important advancements over current software. First, CoCo allows specification of a custom LD Matrix, rather than deriving LD from genotype files. Crucially, this allows the user to derive LD from allelic dosages, rather than hard-called genotypes, which may significantly increase accuracy. Second, the flexible design of our software (ie, the specification of a user-defined LD matrix) and straightforward underlying statistical hypotheses allow this method to be applied beyond GWAS to perform conditional analysis on a wide range of summary statistics, provided that an appropriate correlation matrix is available.

Here, we demonstrate application of our method to perform conditional analyses on eQTL, genetically regulated gene expression (GREX) and phenome-wide association study (PheWAS) data. Application of CoCo therefore allows conditional analyses to be performed using summary statistics from a range of standard genomic and related secondary analyses.

Disclosures: LD and ES are employees of Regeneron Genetics Center

PrgmNr 3173 - Fast and scalable polygenic risk modeling with Variational Inference

View session detail

Author Block: **S. Zabad**[1], S. Gravel[2], Y. Li[1]; [1]McGill Univ., Montreal, QC, Canada, [2]Mc Gill Univ, Montreal, QC, Canada
Disclosure Block: **S. Zabad:** None.
Polygenic risk score (PRS) models are increasingly becoming a valuable diagnostic tool for a variety of complex diseases, providing clinically actionable insights and enabling personalized clinical interventions. Despite their promise, with the exception of a few traits, polygenic scores have not seen wide-scale adoption in clinical practice because their accuracy and robustness remain limited. Many statistical approaches have been developed to extract polygenic scores from biobank-scale datasets. However, statistical challenges abound, illustrated by the limited accuracy of many polygenic scores beyond idealized settings.
A central challenge in this setting is the very high dimensional nature of the data, paired with limited, though rapidly growing, sample sizes. To deal with this challenge, Bayesian approaches, which impose meaningful sparse priors over the effect sizes, have shown great promise. Most of these methods, including SBayesR (Lloyd-Jones et al. 2019) and LDPred (VilhjÃ¡lmsson et al. 2015, Prive et al. 2020), utilize stochastic approximate posterior inference schemes, such as Gibbs sampling. While these methods work well for some polygenic traits, their efficiency and flexibility are hampered by the choice of posterior inference algorithms.
In this work, we present Variational Inference PRS (VIPRS), a Bayesian PRS framework that uses Variational Inference (VI), a deterministic approximate posterior inference scheme. The model is based on the Variational Expectation-Maximization inference algorithm, with closed form updates for the variational parameters as well as hyperparameters, such as the residual variance and the proportion of causal variants. Our model works with individual-level data and summary statistics and can reliably estimate a number of genetic quantities of interest, such as SNP heritability.
We apply VIPRS to simulated and real phenotypes from the White British cohort in the UK Biobank (N = 337195), and demonstrate that it achieves competitive predictive performance with state-of-the-art methods, outperforming SBayesR and LDPred in many settings. We also undertake an ablation study to examine the effects of Linkage Disequilibrium (LD) reference panels and LD estimators on the performance of the method. In our experiments, VIPRS achieves the best predictive performance with non-sparse LD matrices, and our software implementation enables it to run with these large on-disk matrices fairly efficiently and with minimal memory overhead.

PrgmNr 3174 - Fine-mapping robustness in real data

View session detail

Author Block: **R. Cui**[1,2], R. Elzur[1,2], M. Kanai[1,2,3,4], J. C. Ulirsch[1,2,5], B. M. Neale[2,1], M. Daly[1,6], Z. Fan[7], H. K. Finucane[1,2]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [3]PhD program in Bioinformatics and Integrative Genomics, Harvard Med. Sch., Boston, MA, [4]Dept. of Statistical Genetics, Osaka Univ. Graduate Sch. of Med., Osaka, Japan, [5]PhD program in Biological and BioMed. Sci., Harvard Med. Sch., Boston, MA, [6]Inst. for Molecular Med. Finland, Univ. of Helsinki, Helsinki, Finland, [7]Dept. of Statistics and Data Sci., Yale Univ., New Haven, CT

Disclosure Block: **R. Cui:** None.

Fine-mapping algorithms have been useful for narrowing down broad GWAS associations to a smaller set of putative causal variants. Popular fine-mapping methods (ABF, SuSiE and FINEMAP) use a Baysian approach to assign each variant a probability of causality (Posterior Inclusion Probability, PIP). Previously, these methods have been shown to be calibrated in simulations. However, simulations usually don't capture the complexities in real data. We aim to evaluate and improve the robustness and calibration of these methods in real data.

To evaluate robustness in real data, we performed fine-mapping of 10 well-powered quantitative phenotypes on a random subset of 100K individuals of white British ancestry from the UK Biobank and compared these results to existing fine-mapping results at a sample size of 360K. We observed inconsistency in the form of non-replication, where high-PIP (>0.9) variants at 100K can drop to low-PIP (Functional enrichment analyses show that 40Â±8% of non-replicating variants are in coding, regulatory or conserved regions. Similar proportion is annotated for low-PIP variants at 360K when matched on PIP with the non-replicating variants. Whereas 91Â±6% of PIP-matched (at 360K) replicated variants are annotated, suggesting non-replicating variants are likely non-causal, thus indicating miscalibration in real data.

We found that the following properties differed non-trivially between replicated and non-replicating variants: PIP, PIP difference between SuSiE and FINEMAP, and whether or not the variant is in a SuSiE 95% credible set. We also examined: MAF, INFO score, LD score, and number of credible sets within 100kb, etc. We did not find any property that, on its own, could accurately classify replication status.

Our realistic simulations show that prior misspecification can lead to miscalibration. When there is a false positive, it is often easier to detect than the true causal variant it's in LD with. We measure detectability of each variant by computing the probability of obtaining high-PIP using single-causal-model fine-mapping given a certain effect size, prior variances and local LD pattern. Our analysis, however, has not concluded that prior misspecification is the major driver of non-replication or miscalibration in real data.

PrgmNr 3175 - Functional Response Regression Model on Correlated Longitudinal Microbiome Sequencing Data

View session detail

Author Block: **B. Chen**, W. Xu; Princess Margaret Cancer Ctr., Toronto, ON, Canada
Disclosure Block: **B. Chen:** None.

Functional regression has been widely used on longitudinal data, but it is not clear how to apply functional regression to microbiome sequencing data. We propose a novel functional response regression model analyzing correlated longitudinal microbiome sequencing data, which extends the classic functional response regression model only working for independent functional responses. We derive the theory of generalized least squares estimators for predictors' effects when functional responses are correlated, and develop a data transformation technique to solve the computational challenge for analyzing correlated functional response data using existing functional regression method. We show by extensive simulations that our proposed method provides unbiased estimations for predictors' effect, and our model has accurate type I error and power performance for correlated functional response data, compared with classic functional response regression model. Finally we implement our method to a real infant gut microbiome study to evaluate the relationship of clinical factors to predominant taxa along time.

PrgmNr 3176 - Genealogical models of LD for Bayesian applications in GWAS

View session detail

Author Block: **P. Salehi Nowbandegani**[1], W. Wohns[2,1], A. Bloemendal[1], B. M. Neale[3], L. O'Connor[1]; [1]Broad Inst., Cambridge, MA, [23]Massachusetts Gen. Hosp., Boston, MA
Disclosure Block: **P. Salehi Nowbandegani:** None.
Pervasive linkage disequilibrium (LD) among nearby common SNPs is a primary challenge in statistical applications of GWAS like heritability estimation, risk prediction and fine mapping. LD is highly structured, reflecting genealogical history. Until recently, this history could not be inferred, but with recent breakthroughs, biobank-scale whole-genome genealogies can be reconstructed and encoded as a sequence of trees along the genome (Kelleher et al. 2019). In the tree sequence, alleles are identified with the mutant ancestral haplotype where they were first observed.
We show that statistical relationships (conditional dependencies) among linked SNPs correspond to genealogical relationships among the mutant ancestral haplotypes. Without recombination, the correspondence is straightforward. If allele b arises on a haplotype carrying allele a, and allele c arises on an ab haplotype, the conditional LD structure is c-b-a: although a and c may be in LD, they are conditionally independent given b. These conditional relationships are encoded in the LD graphical model. We developed a method to compute the LD graphical model from an inferred tree sequence (with recombination), showing that it contains all conditional dependencies as connected sets.
LD graphical models can be applied to GWAS summary statistics, whose sampling distribution is efficiently described by an LD graphical model with edge weights (i.e. a precision matrix). An immediate application is to fixed-effects heritability estimation (Shi et al. 2017). The sparsity of our inferred LD precision matrix is especially useful in Bayesian applications, where conditional independencies enable efficient algorithms. We derive a fast variational algorithm, which can be applied to polygenic risk prediction and fine mapping. This approach extends to association data from multiple traits and multiple ancestry groups.

PrgmNr 3177 - Genetic association analysis of a binary trait detects more than just the genetic effect: implications for pleiotropy and replication studies

View session detail

Author Block: **Z. Zhang**, L. Sun; Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **Z. Zhang:** None.
Pleiotropy analyses of multiple phenotypes are ubiquitous, and summary statistics such as the effect size estimate, test statistic are the building blocks for such analyses.While it is straightforward to aggregate summary statistics derived from analyzing multiple continuous traits, we show that a critical problem arises when analyzing binary traits. That is, power of testing the association between a genetic variant G and a binary trait Y *also* depends on the effect of other covariates such as age and sex, which may vary between the traits of interests.To demonstrate this, consider a simple logistic regression model, logit(Prob(Y=1))=beta.0+beta.G*G+beta.E*E. We show analytically that power of testing the null hypothesis, H0: beta.G=0, inversely depends on the magnitude of beta.E, in addition to beta.G, sample size and the minor allele frequency (MAF) of the variant of interest. This is not the case when analyzing a continuous trait using a Gaussian linear model.We confirm our theoretical results by simulation. We assume two binary traits measured in the same set of n=1000 individuals, and MAF=0.3 and beta.0=0.5 without loss of generality. We then let beta.G=0.3 for both traits, but beta.E=0.8 for trait Y1 while beta.E=0.3 for trait Y2. The empirical power of testing H0: beta.G=0 is 62% for Y1 while 86% for Y2. In comparison, if the traits are continuous (with an error term of N(0, 3)), the power of testing H0: beta.G=0 is 52% for both traits, not influenced by the differential covariate effect, beta.E.Our findings have important implications for the current pleiotropy studies of binary traits, and planning of a successful replication study, because the presumed genetic association evidence is relative to other covariate effects, which very likely vary between different phenotypes, populations, or samples collected using different ascertainment schemes.

PrgmNr 3178 - GenoToolKit: A convenient and accessible R package for genetic data preparation and analysis

View session detail

Author Block: **D. Loesch**, T. D. O'Connor; Univ. of Maryland Sch. of Med., Baltimore, MD
Disclosure Block: **D. Loesch:** None.

Data preparation is a significant part of statistical genetic workflows in terms of time consumption as well as importance. It is critical to have a unified ecosystem for data preparation, both for reproducibility and for efficiency. Here, we present a convenient and accessible R package, genoToolKit, that can fit into current workflows and stand alone as an analysis toolkit. Functionality of the genoToolKit package is split into three main focus areas: data preparation, statistical genetics, and scans for selection, with the primary emphasis on data preparation. In the case of data preparation, the primary function is alignVCF(). This function aligns a genotype file against a reference for imputation or merging, handling strand flips and allele switches. The package also contains a number of functions for the manipulation of genetic data both within R and via convenient wrapper functions for PLINK and BCFtools. Using these functions, we were able to efficiently and uniformly prepare data for 18,608 subjects for a polygenic risk score (PRS) study as part of the Genetics of Latin American Diversity (GLAD) database project. For the second focus area, the package contains functions for performing an admixture mapping analysis, calculating a PRS, evaluating the PRS using cross-fold validation, and a wrapper function for performing GWAS with PLINK. There are also several related functions for preparing data as input for popular PRS software packages. For the third area of focus, there are functions for performing scans for selection using the population branch statistic and by detecting enrichment of a particular ancestry at a given locus. Overall, this R package can assist with data preparation and perform select analyses in a convenient-to-use manner. To support its usage from the command line, we have created resources for building a Docker image with argument handling (https://github.com/dloesch/genoToolKit-docker). The R package is available at https://github.com/dloesch/genoToolKit.

PrgmNr 3180 - Novel Bayesian TWAS method using eQTL summary statistics

View session detail

Author Block: **Q. Dai**[1], L. Franke[2], U. VÃµsa[3], G. C. Gibson[4], M. P. Epstein[5], J. Yang[5]; [1]Dept. of Biostatistics and Bioinformatics, Emory Univ., Atlanta, GA, [2]Univ. Med. Ctr. Groningen, Groningen, Groningen, Netherlands, [3]Univ. Med. Ctr. Groningen, Groningen, Groningen, Netherlands, [4]Georgia Tech, Atlanta, GA, [5]Emory Univ, Atlanta, GA
Disclosure Block: **Q. Dai:** None.
Transcriptome-wide association studies (TWAS) have been widely used to identify gene-trait associations. Existing TWAS tools like PrediXcan and TIGAR first estimate expression quantitative trait loci (eQTL) effect sizes using a regression model with expression trait as the response variable and cis-SNP genotypes as predictors, and then take these eQTL effect size estimates as SNP weights for a gene-based association test. However, such regression models require access to a reference panel with individual-level genetic and transcriptomic data. Thus, existing methods are unable to utilize public eQTL summary statistics published in recent meta-analyses from large-scale consortiums such as the eQTLGen consortium of blood tissues.

To fill this important gap, we propose a TWAS method that allows for using only eQTL summary statistics as reference. Our method employs a summary-statistics based Bayesian regression method originally proposed for polygenic risk score analyses (PRS-CS), which places a continuous shrinkage prior on cis-eQTL effect sizes to shrink small effect sizes of potentially false eQTL towards zero. Our method can infer posterior effect sizes of cis-eQTLs using only eQTL summary statistics from standard eQTL analyses based on single variant tests and an external linkage disequilibrium (LD) reference panel. These estimated cis-eQTL effect sizes can be used for TWASs using either individual-level or summary-level GWAS data. We develop a novel tool using Python and Bash scripts to implement this Bayesian regression model for TWAS, which is referred to as BS-TWAS. Simulation studies using real genotype data from the Religious Orders Study and Rush Memory Aging Project and simulated expression data illustrated that our BS-TWAS method achieved comparable results with PrediXcan and TIGAR using individual-level reference data, with respect to both gene expression prediction R^2 and TWAS power. We trained gene expression prediction models by BS-TWAS for 1777 genes on chromosomes 2 and 3 with eQTL summary statistics from the eQTLGen consortium (n = 31,684) and the LD reference from GTEx V8 samples, and then tested the prediction accuracy using a non-overlapping set of 315 blood samples from the GTEx V8. We found that 48% of the genes have an imputation R^2 >0.01 with median 0.0335 and range (0.01, 0.36).

These preliminary results showed that BS-TWAS using only eQTL summary statistics obtained comparable results as using individual-level reference data, indicating the effectiveness of using large-scale eQTL summary statistics to unravel gene-trait associations by TWAS.

PrgmNr 3181 - Polygenic risk prediction using gradient boosted trees captures non-linear genetic effects and allele interactions in complex phenotypes

View session detail

Author Block: **T. Sofer**[1], M. Elgart[1], G. Lyons[2], S. Romero-Brufau[2], N. Kurniansyah[3], J. A. Brody[4], X. Guo[5], H. J. Lin[6], L. M. Raffield[7], Y. Gao[8], H. Chen[9], P. S. de Vries[9], D. Lloyd-Jones[10], L. A. Lange[11], G. M. Peloso[12], M. Fornage[13], J. I. Rotter[14], S. S. Rich[15], A. C. Morrison[16], B. Psaty[17], D. Levy[18], S. Redline[1], The NHLBIâ□□s Trans-Omics in Precision Medicine (TOPMed) Consortium; [1]Harvard Med. Sch., Boston, MA, [2]Harvard Chan Sch. of Publ. Hlth., Boston, MA, [3]Brigham and Women's Hosp., Boston, MA, [4]Univ of Washington, Seattle, WA, [5]Harbor-UCLA, Torrance, CA, [6]Harbor-UCLA Med. Ctr., Palos Verdes Estates, CA, [7]UNC - Chapel Hill, Chapel Hill, NC, [8]Univ. of Mississippi Med. Ctr., Jackson, MS, [9]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [10]Northwestern Univ., Chicago, IL, [11]Univ Colorado Denver, Aurora, CO, [12]Boston Univ., Boston, MA, [13]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [14]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [15]Univ Virginia, Charlottesville, VA, [16]Univ. of Texas at Houston, Houston, TX, [17]Univ. of Washington, Seattle, WA, [18]NHLBI/NIH, Framingham, MA
Disclosure Block: **T. Sofer:** None.

Polygenic risk scores (PRSs) are commonly used to quantify the inherited susceptibility for a given trait. However, the scores fail to account for non-linear and interaction effects between single nucleotide polymorphisms (SNPs). Machine learning algorithms can be used to account for such non-linearities and interactions. We trained and validated polygenic prediction models for five complex phenotypes in a multi-ethnic population: total cholesterol, triglycerides, systolic blood pressure, sleep duration, and height. We used an ensemble method of LASSO for feature selection and gradient boosted trees (XGBoost) for non-linearities and interaction effects. In an independent test set, we found that combining a standard PRS within a XGBoost model increases the percentage of variance explained (PVE) of the PRS by: 25% for sleep duration, 26% for height, 44% for systolic blood pressure, 64% for triglycerides, and 85% for total cholesterol. Machine learning models trained in specific racial/ethnic groups performed similarly in multi-ethnic trained models, despite smaller sample sizes. The predictions of the machine learning models were superior to the PRS in each of the racial/ethnic groups in our study. However, among Blacks the PVE was substantially lower than for other groups. For example, the PVE for total cholesterol was 8.1%, 12.9%, and 17.4% for Blacks, Whites, and Hispanics/Latinos, respectively. This work demonstrates an effective method to account for non-linearities and interaction effects in genetics-based prediction models.

PrgmNr 3182 - Polygenic transcriptome risk scores improve cross-ethnic portability for COPD and lung function in the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program

View session detail

Author Block: **X. Hu**[1], M. H. Cho[2], H. Im[3], A. W. Manichaikul[1], TOPMed Lung Working Group; [1]Ctr. for Publ. Hlth.Genomics, Univ. of Virginia, Charlottesville, VA, [2]Channing Div. of Network Med., Dept. of Med., Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [3]The Univ. of Chicago, Chicago, IL

Disclosure Block: **X. Hu:** None.

Background: Chronic Obstructive Pulmonary Disease (COPD), diagnosed by reduced lung function, is a highly heterogeneous disease. Existing polygenic risk scores (PRS) enable early identification of genetic risk for COPD. However, predictive performance of the PRS is limited when the discovery and target populations are not well matched. **Methods**: To improve cross-ethnic portability of risk prediction, we introduced a PrediXcan-derived polygenic transcriptome risk scores (PTRS), developed under the hypothesis that the biological mechanisms of disease are shared across ancestry groups. We constructed the PTRS using summary statistics from application of PrediXcan on large scale GWAS of lung function (Forced Expiratory Volume in 1 second (FEV1) and its ratio to Forced Vital Capacity (FEV1/FVC)) from the UK Biobank, which representing primarily European ancestry-based cohort. To examine prediction performance and cross-ethnic portability of the proposed PTRS candidates, we performed smoking-stratified analyses on multi-ethnic training data for 29,381 participants from TOPMed population/family-based cohorts (NHW=14,727, AA=7,025, HIS=7,629). The best risk score candidates were then tested for 11,771 multi-ethnic participants from TOPMed COPD-enriched studies (NHW=8,144, AA=3,627). Analyses were carried out for two dichotomous traits of COPD (Moderate-to-Severe and Severe COPD) and two quantitative lung function traits (FEV1 and FEV1/FVC). **Results**: While the novel PTRS had lower prediction accuracy for European ancestry participants than PRS, the PTRS performed slightly better than PRS for predicting COPD in heavy smoking African Americans (OR=1.24 [95%CI: 1.08-1.43] from PTRS and OR=1.10 [95% CI: 0.96-1.26] from PRS for Moderate-to-Severe COPD; for Severe COPD, OR=1.51 [95%CI: 1.04-12.19] from PTRS and OR=1.31 [95% CI: 0.87-1.96] from PRS). In addition, as hypothesized, the cross-ethnic portability was significantly higher for PTRS than for PRS (pConclusions: Our study demonstrates the value of PTRS for improved prediction of COPD risk in African Americans. Future work will strengthen the PTRS framework by leveraging multi-ethnic gene expression reference data of larger sample sizes and from disease relevant tissues.

PrgmNr 3183 - Predicting overweight and obese with genome-wide, epigenome-wide gene-gene and gene-diet interactions using machine learning

View session detail

Author Block: **Y-C. Lee**[1], J. J. Christensen[2], L. D. Parnell[1], C. E. Smith[3], N. M. McKeown[4], J. M. OrdovÃis[3], C-Q. Lai[1]; [1]USDA ARS, JM-USDA Human Nutrition Res. Ctr. on Aging at Tufts Univ., Boston, MA, [2]Norwegian Natl. Advisory Unit on FH, Oslo Univ. Hosp.; Dept. of Nutrition, Univ. of Oslo, Oslo, Norway, [3]Nutrition and Genomics Lab., JM-USDA Human Nutrition Res. Ctr. on Aging at Tufts Univ., Boston, MA, [4]Nutritional Epidemiology Lab., JM-USDA Human Nutrition Res. Ctr. on Aging at Tufts Univ., Boston, MA

Disclosure Block: **Y. Lee:** None.

Obesity is a main risk factor for many chronic diseases and health conditions. Studies have shown that genetic, epigenetic and environmental factors and their complex interactions contribute to obesity. This study aimed to identify genetic, epigenetic and dietary factors, and their interactions that contribute to overweight and obesity in order to develop precise prevention and treatment strategies. We conducted a combined genome-wide and epigenome-wide scan for main effects and up to three-way interactions among 422,793 single nucleotide polymorphisms (SNPs), 415,202 DNA methylation sites and 397 dietary and lifestyle factors using the Generalized Multifactor Dimensionality Reduction (GMDR) method in the training samples (n=1,573) of the Framingham Heart Study Offspring cohort Exam 8. Using identified genetic, epigenetic and dietary factors, we then applied machine learning (ML) algorithms to predict participantsâ obesity status in the testing samples (n=394). The quality of prediction models was evaluated using Area Under the Receiver Operating Characteristic Curve (ROC-AUC). The GMDR method identified 220 SNPs, 536 epigenetic markers and 50 dietary and lifestyle factors that were highly informative predictors predictive for obesity. Comparing several ML algorithms, we found that the stochastic gradient boosting model provided the best prediction accuracy for obesity in the training set, and overall accuracy of 71.8% and ROC-AUC of 0.70 in the test set samples. Top predictors of the best-fit model were 21 SNPs, 230 epigenetic markers in genes of *CPT1A*, *FSD2*, *ABCG1*, *SLC7A11*, *RNF145*, *SREBF1* and others and 26 dietary factors, including processed meat, diet soda, French fries, high fat dairy, calcium, flavonols, sugar, artificial sweeteners, and alcohol intake. In conclusion, our study has demonstrated a novel method by which integrating genomic, epigenomic gene-gene and gene-environment interactions and applying ML supports prediction of obesity status. This genome-wide approach improves our understanding of obesity by identifying dietary and lifestyle factors that interact with genotype and DNA methylation sites. Results such as these can inform personalized nutrition strategies for the prevention and treatment of obesity.

PrgmNr 3184 - Prioritizing disease genes by integrating GWAS and WES with gene-level functional and perturbation data

View session detail

Author Block: **K. K. Kumar Dey**[1], K. Jagadeesh[2], W. Zhou[3], W. Bi[4], Z. Zhao[5], E. Weeks[6], K. Geiger-Schuller[7], X. Jin[6], R. Xavier[6], S. Lee[8], A. Regev[9,7], H. K. Finucane[10], A. L. Price[11]; [1]Harvard Sch. of Publ. Hlth., BOSTON, MA, [2]Allston, MA, [3]Massachusetts general Hosp., Broad Inst., Boston, MA, [4]1415 Washington HTS, Ann Arbor, MI, [5]Univ. of Michigan, Superior Twp, MI, [6]Broad Inst., Cambridge, MA, [7]Genentech, South San Francisco, CA, [8]Univ MICHIGAN, Ann Arbor, MI, [9]Broad Inst. of MIT and Harvard, Cambridge, MA, [10]Harvard T. H. Chan Sch. of Publ. Hlth., Cambridge, MA, [11]Harvard Sch Pub Hlth, Boston, MA

Disclosure Block: **K.K. Kumar Dey:** None.

GWAS have identified thousands of disease-associated variants, but identifying genes and gene programs that impact disease remains challenging. Here, we propose a new method, PolyGene, that assigns a disease association score to each gene using a polygenic ensemble approach that integrates GWAS data for common variants (MAGMA: de Leeuw et al. 2015 PLoS Comp. Biol, PoPS: Weeks et al. 2020 medRxiv), WES data for rare variants when available (SAIGE-GENE: Zhou et al 2020 Nat. Genet), and gene interactions from protein-protein and co-expression networks.

We applied PolyGene to GWAS summary statistics for 28 diseases and traits (average N=270K) for which gold-standard gene sets were available based on approved drug targets, Mendelian genes and CRISPR screen result implicated genes. PolyGene attained higher excess overlap across these gold-standard gene sets than other gene scoring methods: 5.0x (P=8e-32) higher than MAGMA and 1.4x (P=4e-18) higher than PoPS. For a subset of 7 well-powered UK Biobank blood cell traits (average N=443K), incorporating rare-variant driven gene-based association statistics from WES data (PolyGene-S) produced a small but statistically significant further improvement (1.1x, P=1e-05).

We applied PolyGene to score genes in three gene programs (small gene sets defined using functional data). First, we analyzed post-stimulation perturbation programs for 24 genes in dendritic cells, derived from Perturb-seq data (Dixit et al. 2016 Cell). The perturbation programs for known inflammatory genes such as E2f1 and Nfkb1 attained the highest excess overlap with PolyGene scores across 9 autoimmune diseases (1.5x-2.5x, P=1e-04-4e-09). Second, we analyzed perturbation programs for 35 de novo autism risk variant genes in 5 brain cell types, derived from in-vivo Perturb-seq data (Jin et al. 2020 Science). The perturbation programs in astroglia and microglia attained the highest excess overlap with PolyGene scores for autism (mean of 2.4x, vs. mean of 1.2x in other cell types) with perturbation program for Cul3 gene in microglia demonstrating the highest excess overlaps (4.0x, P=3e-13). Third, we defined autism disease progression programs for 16 brain cell types based on genes that were specifically expressed in autism cases vs. controls in post-mortem brain snRNA-seq data (105K nuclei; Velmeshev et al 2019 Science). The disease progression programs in excitatory neurons and microglia attained the highest excess overlap with PolyGene scores for autism, consistent with perturbation program analysis.

In conclusion, PolyGene is a powerful approach for scoring genes and gene programs for association to disease to produce biological insights.

PrgmNr 3185 - Rare variant effect sizes are associated with variant pathogenicity

View session detail

Author Block: **V. Morrill**[1,2], J. L. Halford[3,1], S. Choi[1], S. J. Jurgens[1], G. Melloni[4], N. A. Marston[4], L-C. Weng[2,1], V. Nauffal[1], A. A. Hall[2,1], S. Gunn[5], C. Austin-Tse[2], J. Pirruccello[2], S. Khurshid[2], H. L. Rehm[2], E. J. Benjamin[6], C. Kooperberg[7], B. M. Psaty[8], S. Redline[3], J. I. Rotter[9], N. Gupta[1], S. Harris[2], E. Wong[2], K. Rice[10], P. Schnatz[11], NHLBI Trans-Omics for Precision Medicine (TOPMed)Consortium, TOPMed EKG Working Group, M. S. Sabatine[4], C. Ruff[4], K. L. Lunetta[5], P. T. Ellinor[2,1], S. Lubitz[2,1]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Harvard Med. Sch., Boston, MA, [4]Brigham and Womenâ□□s Hosp. and Harvard Med. Sch., Boston, MA, [5]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [6]Boston Univ, Framingham, MA, [7]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [8]Univ. of Washington, Seattle, WA, [9]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [10]Univ of Washington, Seattle, WA, [11]The Reading Hosp. of Tower Hlth., Reading, PA

Disclosure Block: **V. Morrill:** None.

Efficient assessment of variant pathogenicity, a genetic susceptibility marker, is critical for research and clinical applications. We tested whether associations between rare variants and quantitative endophenotypes for three monogenic cardiometabolic diseases are proxies for variant pathogenicity. In the UK Biobank, we examined associations between rare coding variants in (1) familial hypercholesterolemia (FH) susceptibility genes and low-density lipoprotein cholesterol (LDL-C) levels (n=189,656 individuals), (2) long QT syndrome (LQTS) genes and electrocardiographic QTc intervals (n=33,521 individuals), and (3) maturity-onset diabetes of the young (MODY) genes and hemoglobin A1c (HbA1c) levels (n=189,744 individuals). We observed strong associations between rare variant effect sizes and pathogenicity assertions in ClinVar (p-values

PrgmNr 3186 - Reproducible analyses of petabytes of Heart Lung Blood and Sleep (HLBS) data in the cloud using NHLBI BioData Catalyst, the data analysis ecosystem for HLBS researchers

View session detail

Author Block: **A. Leaf**[1], A. K. Manning[2], P. Avillach[3], R. BOYLES[4], J. Kaltman[5], B. Paten[6], E. Sheets[7], T. Harris[6], K. Osborn[6], S. Suber[8], I. Borecki[9], NHLBI BioData Catalyst Consortium; [1]Seven Bridges, Charlestown, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Harvard Med. Sch., Boston, MA, [4]RTI Intl., London, United Kingdom, [5]Natl. Heart, Lung, and Blood Inst., Bethesda, MD, [6]UCSC, Santa Cruz, CA, [7]UC Santa Cruz Genomics Inst., Univ. of California Santa Cruz, Santa Cruz, CA, [8]Renaissance Computing Inst., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [9]NHLBI BioData Catalyst Consortium, Bethesda, MD

Disclosure Block: **A. Leaf:** None.

The rate of population health and biomedical data generation is accelerating rapidly. Cloud computing offers an effective way to store and analyze these data, but substantial investment is necessary to make cloud-based research portals broadly usable by the researcher community. Such systems require substantial infrastructure (scalable and secure environments), analysis resources (analysis environments, harmonized data) and training resources for data sharing, collaboration mechanisms, and new statistical methodologies. The National Heart Lung and Blood Institute (NHLBI) initiated the BioData Catalyst effort with a mission to develop and integrate advanced cyberinfrastructure while championing â□□findable, accessible, interoperable, and reusableâ□□ (FAIR) principles to accelerate HLBS research. Now, NHLBI BioData Catalyst provides access to a highly secure, scalable, reproducible, collaborative, and extensible cloud analysis ecosystem with controlled access to more than 3 Petabytes of HLBS data, including harmonized data from NHLBIâ□□s TOPMed Study, BioLINCC, and several NIH COVID-19 datasets. â□□Bring Your Own Dataâ□□ functionality enables users to securely upload and privately analyze datasets on the platform provided that such usage is allowed by their existing data use agreements and IRB policies. Researchers gain immediate access to interactive analysis environments such as R, Python, and SAS. Additionally, researchers can easily deploy their own docker-based tools or configure hundreds of pre-loaded, cost optimized tools and workflows from Dockstore encoded in Common Workflow Language (CWL) or Workflow Description Language (WDL). An open cohort builder allows researchers to quickly evaluate study feasibility and discover data that can accelerate their research. In this presentation, we will show new examples of research being performed on the NHLBI BioData Catalyst, demonstrate the utility of combining datasets to explore new hypotheses, and illustrate how the NHLBI BioData Catalyst, as a community-driven ecosystem, is democratizing data access and advancing HLBS science.

PrgmNr 3187 - Single-cell allele-specific expression analysis and comparison using mixture models

View session detail

Author Block: **G. Qi**[1], B. Strober[1], A. Battle[2]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Johns Hopkins Univ, Baltimore, MD
Disclosure Block: **G. Qi:** None.
Background. Allele specific expression (ASE), which quantifies the imbalance in gene expression between two parental copies, is a powerful tool to study cis-regulatory effects. In particular, differential ASE across conditions, e.g. experimental interventions, cell types, genotypes, or differentiation trajectories can reflect context-specific cis-regulatory effects. Recently, single-cell RNA sequencing (scRNA-seq) has allowed the measurement of ASE at the resolution of individual cells. However, previous statistical methods for identifying differential ASE were developed for bulk RNA-seq, and cannot be directly applied to scRNA-seq due to considerations including sparsity, repeated measurements from the same individual, and sequencing depth. **Methods.** We develop scASEmix, a mixture beta-binomial random-effects model for conducting differential ASE analysis from scRNA-seq data. The model uses a subject specific random effect to account for the correlation among cells from the same individual. In addition, the model conducts implicit phasing between the heterozygous exonic single-nucleotide polymorphism (SNP) and the causal regulatory SNP using latent variables. We estimate the parameters using Expectation-Maximization (EM) algorithm coupled with numerical integration. **Results.** scASEmix accurately captures differential ASE in a wide range of scenarios. We demonstrate through simulation that it has well-controlled type I error, and leads to large power improvements compared to standard beta-binomial models under weak linkage disequilibrium (LD) between the exonic and regulatory SNPs. We use scASEmix to analyze the single-cell ASE data from an endoderm differentiation experiment by Cuomo et al, which was designed to identify dynamic genetic effects on gene expression. We identify novel genes that show significant differential ASE along the differentiation trajectory. These genes have substantial overlaps with previously reported gene sets related to early differentiation. **Conclusions.** We propose a powerful method for differential ASE analysis using scRNA-seq data. The method can be applied to a wide range of scenarios to study context-specific cis-regulatory effects, ultimately illuminating regulatory effects that may underlie processes including cellular differentiation and disease.

PrgmNr 3188 - Sister haplotype association based on recombination disequilibrium

View session detail

Author Block: **H. Xu**; Augusta Univ., Augusta, GA
Disclosure Block: **H. Xu:** None.
Haplotype-based association analysis has several advantages over single-SNP association analysis. However, to date all associations between haplotypes and diseases have not excluded recombination interference among multiple SNP loci within haplotypes and hence some results might be confounded by recombination interference. We developed a new haplotype association method based on recombination disequilibrium (RD). Applying this method to a SNP haplotype dataset of 210 Alzheimer's disease (AD) cases and 159 nondemented elderly controls, we successfully found that some pairs of sister-haplotypes containing ApoE gene were associated with risk for AD under no RD but all those without ApoE-Îµ3 and ApoE-Îµ4 were not associated with risk for the disease and sister-haplotype pairs within genes IL-13 and COMT were not associated with risk for breast cancer. In addition, none of sister-haplotype pairs in IL-17A gene was detected to be associated with risk for coronary artery disease. All the previously reported associations of haplotypes within these genes with risk for these diseases might be due to strong RD and/or inappropriate haplotype pairs.

PrgmNr 3189 - SUMMIT: An integrative approach for better transcriptomic data imputation improves causal gene identification

View session detail

Author Block: **Z. Zhang**[1], J. R. Bradley[1], L. Wu[2], C. Wu[1]; [1]Dept. of Statistics, Florida State Univ., Tallahassee, FL, [2]Univ. of Hawaii Cancer Ctr., Honolulu, HI
Disclosure Block: **Z. Zhang:** None.
Transcriptome-wide association studies (TWASs), which integrate expression reference panels with genome-wide association study (GWAS) results to discover gene-trait associations, have deepened our understanding of genetic regulation in many complex traits. However, the number of analyzable genes and thus the power of TWAS is largely determined by the size of expression reference panels. One obvious but administratively onerous approach is to combine individual-level data from several consortia to increase the sample size. However, privacy concerns and sample consents often preclude access to individual-level genetic data, making a pooled individual-level expression panel unavailable. In this work, we introduce SUMMIT, a novel method that makes it possible to integrate a summary-level expression reference panel with a much larger sample size into GWAS to identify associated genes. In brief, we build gene expression prediction models in blood based on the summary data released by the eQTLGen consortium, which is to date the largest meta-analysis in 31,684 blood samples from 37 cohorts. Compared with benchmark methods, MR-JTI, TWAS-Fusion, PrediXcan, and UTMOST, SUMMIT built more gene expression prediction models (10,026 with R^2> 0.01) and achieved significantly higher prediction accuracy in different quantiles (p -9 by K-S test). To evaluate the performance of identifying significant associations, we applied SUMMIT to the summary statistics from 25 GWAS. Compared with benchmark methods, SUMMIT identified substantially more associations for each trait analyzed, showing 222% improvement compared with MR-JTI ($p=3.1*10^{-9}$ by the Wilcoxon rank test), 306% improvement compared with TWAS-fusion, 264% improvement compared with PrediXcan, and 211% improvement compared with UTMOST. Next, we compared different methods in terms of identifying the likely causal genes that mediate the associations between GWAS loci and the traits of interest by using a set of 1,424 likely causal gene-trait pairs curated by using the OMIM. We show that the SUMMIT yielded good sensitivity and specificity for identifying the silver standard genes and achieved the highest AUC (0.701) among all the methods compared (MR-JTI 0.6329; PrediXcan 0.6164; TWAS-fusion 0.6041; UTMOST 0.6213). More importantly, SUMMIT was applicable to analyze genes with smaller heritable expressions (0.005 2

PrgmNr 3192 - TIGAR-V2: Efficient TWAS Tool with Nonparametric Bayesian eQTL Weights of 49 Tissue Types from GTEx V8

View session detail

Author Block: **J. Yang**[1], R. Parrish[2], G. Gibson[3], M. P. Epstein[1]; [1]Emory Univ, Atlanta, GA, [2]Emory Univ., Decatur, GA, [3]Georgia Inst. of Technology, Atlanta, GA
Disclosure Block: **J. Yang:** None.
Standard Transcriptome-wide association study (TWAS) methods first train gene expression prediction models using reference transcriptomic data, and then test the association between the predicted genetically regulated gene expression (GReX) and phenotype of interest with these trained models and GWAS data. We previously developed a TWAS tool called **Transcriptome-Integrated Genetic Association Resource (TIGAR)** that can train gene expression imputation models using either nonparametric Bayesian Dirichlet Process Regression (DPR) or Elastic-Net penalized regression (as used by PrediXcan). TIGAR can perform TWAS using either individual-level or summary-level GWAS data. Besides the Burden type TWAS test, the software further implements an additional variance-component test for TWAS that retains power under model misspecification.
Most existing TWAS tools like PrediXcan often require an extra step of preparing genotype input files from the commonly used VCF format, and extra coding to enable parallel computation. To improve the efficiency of TWAS tools to train gene expression prediction models using oneâ□□s own reference transcriptomic data, we develop an improved version of the TIGAR tool (**TIGAR-V2**), which directly reads VCF files, enables parallel computation, and reduces up to 90% computation cost compared to the initial version. Using **TIGAR-V2**, we trained nonparametric Bayesian DPR gene expression prediction models for 49 tissues from the Genotype-Tissue Expression (GTEx) project V8 reference data.
We illustrated the usefulness of these DPR eQTL weights through TWAS of breast and ovarian cancer utilizing public GWAS summary statistics. Using DPR weights, we identified 88 and 37 risk genes respectively for breast and ovarian cancer, most of which are either known or near previously identified GWAS (~95%) or TWAS (~40%) risk genes of the corresponding phenotype. Moreover, three novel independent risk genes of breast cancer were identified with known functions in carcinogenesis, and two of these are shared by ovarian cancer. Using PrediXcan weights estimated by Elastic-Net, we identified only 56 and 4 risk genes respectively for breast and ovarian cancer (37.5% of these breast cancer risk genes and all ovarian risk genes were also identified by using DPR weights) and missed these three novel findings. These findings suggest that using DPR weights can provide additional insight into the transcriptional regulation of complex human traits.
Our **TIGAR-V2** TWAS tool with the trained nonparametric Bayesian DPR eQTL weights of 49 tissues from GTEX V8 will be freely available on Github, providing a useful resource for mapping risk genes of complex diseases.

PrgmNr 3194 - Trans-Ancestry Fine-Mapping using 3.4 Million Individuals from diverse ancestries elucidates the genetic architecture of tobacco and alcohol use and addiction

View session detail

Author Block: **X. Wang**, GWAS & Sequencing Consortium of Alcohol and Nicotine, Trans-Omics for Precision Medicine; Penn State Coll. of Med., Hershey, PA
Disclosure Block: **X. Wang:** None.
Tobacco and alcohol use are heritable traits and leading causes for many diseases. Recently, breakthrough in addiction genetics has been made through GWAS meta-analysis in large samples of European ancestry, yet, the genetic architecture in non-European ancestry remains under-studied. To address this, we have aggregated datasets from 60 cohorts with a total of 3.4 million individuals from diverse ancestries (2,669,029 European, 298,624 East Asian, 285,155 Latinx/Hispanic American, 121,858 African/African American). We fine-mapped 2,543 distinct genomic regions defined by LD structure around conditionally independent variants, the vast majority of which are novel discoveries. This dataset offers an unprecedented opportunity to advance our understanding on the genetic architecture for smoking and drinking behaviors in global populations. We propose an improved meta regression-based model for trans-ancestry genetic effect distributions. Specifically, we use the principal components (PCs) of genome-wide allele frequencies as proxies of continuously varying cohort-level ancestry. We model the genetic effect from each study as a mixture of models with different number of PCs, which can encompass different extent of heterogeneity for each variant. For example, model with 0 PC supports homogenous effects. As the 1st PC separates European and Asian ancestry, model with 1 PC can be interpreted as having heterogenous effects along the European-Asian cline. By imposing a Dirichlet-Multinomial prior, we borrow strength across variants, learn the genetic architecture and fine map causal variants. We perform simulations across different scenarios that assume variants have homogenous effects and/or ancestry-specific effects. We show that our method greatly improves fine mapping resolution and allows us to estimate the fractions of loci that show homogenous effects and ancestry-specific effects. We apply our method to the GSCAN study of 4 smoking phenotypes, i.e., age of initiation of smoking (AgeSmk), cigarettes per day (CigDay), smoking initiation (SmkInit) and smoking cessation (SmkCes), and 1 drinking phenotype, drinks per week (DrkWk). Among 2,543 identified significant loci with a median of 3,274 variants per locus, our proposed method fine-map 34.5% of them to less than 6 variants and 1.51 genes in 90% credible sets, a significant improvement over fine mapping using European ancestry only. Our new results and continued research will elucidate the genetic architecture in global ancestries.

PrgmNr 3195 - Using (Natural) Direct and Indirect Effects of Intermediate Risk Factors between Genotype and Disease to Inform Polygenic Risk Scores

View session detail

Author Block: **M. Vujkovic**[1], R. Daniel[2], R. Kember[1], N. Tsao[1], V. Walker[3], J. P. Casas[4], T. L. Assimes[5], Y. Sun[6], K. Cho[7], P. W. Wilson[8], J. M. Gaziano[9], C. J. Oâ□□Donnell[10], P. TSAO[11], D. J. Rader[12], B. F. Voight[1], K-M. Chang[1], S. M. Damrauer[13]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Cardiff Univ., Cardiff, United Kingdom, [3]venexia.walker@bristol.ac.uk, University of Bristol, United Kingdom, [4]VA Boston Hlth.care, [5]Stanford Univ Sch Med., Stanford, CA, [6]Emory Univ, Atlanta, GA, [7]VA Boston Hlth.care System, Boston, MA, [8]Emory Univ. Hosp., Atlanta, GA, [9]VA Boston Hlth.care System, Roxbury Crossing, MA, [10]the NHLBIâ□□s Framingham Heart Study, Framingham, MA, [11]VAPAHCS, Palo Alto, CA, [12]Univ of Pennsylvania, Philadelphia, PA, [13]Hosp. of the Univ. of Pennsylvania, Philadelphia, PA

Disclosure Block: **M. Vujkovic:** None.

In human genetics, there is increasing interest into the underlying mechanisms that explain the effect of genotype on outcome. Within the spectrum of cardiometabolic traits such as type 2 diabetes and cardiovascular disease, there are numerous intermediate risk factors between genotype and disease such as body weight, total cholesterol, LDL-C, glucose, HbA1c, blood pressure, smoking, and alcohol use. A substantial number of genetic variants have been found to be associated with multiple traits (e.g. pleiotropy), however it remains unclear whether they exert independent effects on multiple traits (e.g. horizontal pleiotropy) or possibly act in a mediated fashion (e.g. vertical pleiotropy). Our aims are twofold: first, to identify genetic variants exhibiting full or partial vertical pleiotropy through an intermediate risk factor by use of formal mediation analysis and decompose the total effect of genotype on disease into a (natural) direct and indirect effect. Second, to assess whether the replacement of total effect size in the conventional â□□sum-of-weighted-allelesâ□□ PRS model with direct and indirect effect weights ultimately improves disease prediction. As a proof of concept we test a single mediator model, with obesity as the mediator and type 2 diabetes as the outcome. We applied genome-wide mediation analysis by using the â□□product-of-coefficientsâ□□ regression method on summary statistics of BMI from GIANT+UKBB and T2D from DIAMANTE Consortium. In addition, we performed mediation analysis by means of the causal counterfactual framework with individual level data in 69,869 T2D cases and 127,197 controls in the Million Veteran Program restricted to a subset of 1,101 cardiometabolic trait-associated SNPs. As a result, the two methods showed remarkable concordance of quantifying vertical pleiotropy, and identified known loci that have strong mediating effects on type 2 diabetes through BMI, of which FTO, TCF7L2, IRS1, PPARG, JAZF1, and IGF2BP1 have been established previously. In particular, the TCF7L2 locus additionally exhibits interaction, meaning that within strata of BMI the size of the indirect effect on type 2 diabetes varies. Finally, a mediator-adjusted PRS outperformed the conventional PRS (e.g. sum of weighted alleles) method in predicting T2D (AUC 0.69 vs 0.64). To conclude, we show that a risk model that more precisely models the complex relationship between genotype and disease by including preceding risk factors as intermediates more accurately predicts T2D. It is anticipated that this methodology will take us one step closer towards utilizing genetic variants in clinical risk prediction models.

PrgmNr 3196 - Allele-specific expression of SNPs involved in the immune response and gene transcription is associated with BMI

View session detail

Author Block: **A. Keshawarz**[1], T. Huan[2], R. Joehanes[1], C. Yao[3], C-T. Liu[4], C. Demirkale[5], V. Ramachandran[6], L. A. Cupples[7], D. Levy[8]; [1]Natl. Heart, Lung, and Blood Inst., Framingham, MA, [2]Univ. of Massachusetts Med. Sch., Worcester, MA, [3]Framingham Heart Study, Framingham, MA, [4]Boston Univ. SPH, Boston, MA, [5]NIH, Bethesda, MD, [6]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [7]Boston Univ Sch Pub Hlth, Boston, MA, [8]NHLBI/NIH, Framingham, MA
Disclosure Block: **A. Keshawarz:** None.
Background. Body mass index (BMI) is an estimate of general adiposity associated with numerous clinical outcomes, including subclinical and clinical cardiovascular and metabolic disease. Growing genome-wide association studies have identified hundreds of BMI-associated loci, but the transcriptomic signature of BMI is incompletely understood. The objective of this study was to quantify the association between allelic imbalance and BMI using allele-specific expression analysis in conjunction with GWAS of BMI. **Methods.** Whole genome sequencing and RNA sequencing data were collected from 720 Framingham Heart Study (FHS) Offspring participants (59% women, mean age 66Â±8 years) and 954 FHS Third Generation participants (52% women, mean age 46Â±9 years) as part of the Trans-Omics for Precision Medicine (TOPMed) program. Heterozygous SNPs (n=780,599) were evaluated for allele-specific expression, and the ratio of reference allele to total allele counts for SNPs was calculated. SNPs with significant allelic imbalance based on a Bonferroni-corrected p-value (0.05/780,599) in a binomial test were subsequently evaluated in multiple linear regression to test the association between the reference allele/total allele count ratio and BMI after adjustment for age, sex, and family structure. SNPs significantly associated with BMI at FDR cis-expression quantitative trait loci (eQTLs) to investigate other known functions and associations. **Results.** After adjustment for age, sex, and family structure, allelic imbalance in 139 SNPs was significantly associated with BMI. SNPs identified were annotated to genes that were enriched for cell activation and immune response pathways, and the top five most significant SNPs identified were annotated to *FLYWCH1, ZDHHC6, MYADM*, and *TCL1A*. Twenty-seven SNPs were in regulatory regions; of these, 18 were in promoter or promoter flanking regions. Seven SNPs identified in these analyses were associated with traits in the GWAS catalog, including BMI. Furthermore, these SNPs overlapped with 181 unique *cis*-eQTLs (pPLGLB2, HLA-J, TRIM31, DHDDS, AL671883.2, DPY19L1P2, AL022345.4). **Conclusion.** SNPs implicated in the immune response pathway and in gene regulation show significant allelic imbalance associated with BMI; thus, allele-specific expression may partially explain population-level variation in BMI.

PrgmNr 3197 - Causal gene-to-trait effect estimation for schizophrenia and bipolar disorder using MRLocus

View session detail

Author Block: **M. Love**, N. Matoba, J. L. Stein; UNC-Chapel Hill, Chapel Hill, NC
Disclosure Block: **M. Love:** None.
A number of integrative statistical methods, including TWAS, allow for gene-level aggregation of association signals from individual genetic variants, leveraging tissue-specific eQTL and GWAS study summary statistics. TWAS has successfully been applied to schizophrenia and bipolar disorder GWAS to identify candidate genes implicated in these disorders. However, genes identified as significant by TWAS may not reflect a causal relationship between gene expression and a downstream trait. We recently published a statistical method, MRLocus, that leverages allelic heterogeneity of eQTL signal for a candidate gene, to prioritize genes causally involved in trait variation. Consistent and proportionate effects across multiple eSNPs provides evidence beyond existing TWAS approaches that gene expression levels in a particular tissue mediate the heritability of the downstream trait, and allows for estimation of the gene-to-trait effect. MRLocus gave more accurate estimates of these effects as compared to other cis-MR approaches.
Here, we apply MRLocus to eQTL data from adult post-mortem brain from PsychENCODE (1,387 brain samples) and GWAS of schizophrenia (PardiÃ±as et al, 2018, PGC, 40,675 cases) and bipolar disorder (Stahl et al, 2019, PGC, 20,352 cases), in order to further prioritize genes implicated in previous TWAS-significant results (Gandal et al, 2018), and to associate those genes with causal effect sizes. We estimate MR slopes from eQTL effect sizes from standardized expression, and GWAS effect sizes from log odds ratios. We find 28 genes with significant gene-to-trait effects for schizophrenia, including large effect sizes (absolute value of MR slope > 0.3) for DCC, DCLK3, DDHD2, EMB, FAM109B, FTSJ2, GLYCTK, PAK6, and TPRKB (by 95% credible interval). An MR slope of 0.3 implies that modifying gene expression by one SD will increase disorder risk by ~1/3, well above the effect size expected of one individual SNP on disorder risk. We find 3 genes with significant and large gene-to-trait effect sizes for bipolar disorder: DCLK3, HAPLN4, and RP11-182J1.16 (95% CI), as well as large and marginal significant effects for SNAP91 and TTC39A (by 90% credible interval).
We find that MRLocusâ□□s estimation of the causal effect of eQTLs on risk for neuropsychiatric disorders provides useful information for prioritizing experiments to understand how perturbation of gene expression may mediate disorder risk. Gene-to-trait effect sizes are much larger than those of individual SNPs on disorder risk, suggesting that manipulation of gene expression may have substantial therapeutic influences despite the fact the individual GWAS variants have low effect sizes.

PrgmNr 3198 - Common and rare genetic risk for autism have opposing effects on cognition at autism-associated genes

View session detail

Author Block: **D. Weiner**[1], J. Grove[2], J. Fu[3], F. Satterstrom[1], H. Brand[3], iPSYCH Consortium, L. O'Connor[1], M. J. Daly[4], A. D. Borglum[5], M. E. Talkowski[4], E. Robinson[6]; [1]Broad Inst., Cambridge, MA, [2]Aarhus Univ., Aarhus C, Denmark, [3]Boston, MA, [4]Massachusetts Gen. Hosp., Boston, MA, [5]Aarhus Univ, Aarhus, Denmark, [6]

Disclosure Block: **D. Weiner:** None.

Multiple classes of genetic variation contribute risk to autism spectrum disorder (ASD), and integrating across these lines of evidence remains an important goal to clarify underlying ASD pathogenesis. Substantial ASD liability originates from common polygenic variation and rare *de novo* coding variation. Paradoxically, these two classes of genetic risk exert opposing effects on cognition in ASD: common polygenic variation is often associated with higher IQ, and rare variation is associated with decreased IQ. While gene set enrichment methods such as MAGMA and stratified LD-Score regression estimate that common variation is concentrated around genes implicated in ASD through rare variation, these approaches do not specify *how* this subset of common variation affects cognition. In other words, it is unclear whether the common variant enrichment at ASD-associated genes is increasing or decreasing IQ, a critical question for understanding the convergence of common and rare genetic risk in ASD.

To answer this question, we adapted a genome-wide common variant association approach, the polygenic Transmission Disequilibrium Test (pTDT), to estimate parent-to-child transmission of polygenic risk scores made from small regions of the genome (Stratified-pTDT, or S-pTDT). By constructing a polygenic score from SNPs within ASD-associated genes, we are able to test whether ASD cases over-inherit IQ-increasing alleles from this small region of the genome. We first identified an updated set of 255 genes implicated in ASD through exome sequencing. We then defined a stratified polygenic score from all SNPs within the boundaries of these genes (+/- 10kb) from a recent GWAS of educational attainment (EA, which we use as a genetic proxy for cognition given their high genetic correlation and the significantly larger sample size of the educational attainment GWAS). We then applied S-pTDT in two independent cohorts of ASD trios: SFARI (Simons Simplex Collection and SPARK, 5,048 trios) and the Psychiatric Genomics Consortium (4,334 trios).

We estimate that ASD cases over-inherit stratified polygenic liability for EA at ASD-associated genes in both cohorts ($P = 0.02$ in SFARI, $P = 0.007$ in PGC). Restated, ASD cases on average inherit IQ increasing alleles at genes associated with ASD through IQ-decreasing loss-of-function variation. This analysis suggests that the convergence of common variation and rare variation at these ASD-associated genes confers ASD liability through at least partially distinct mechanisms. Our analysis also introduces S-pTDT, a novel statistical approach for associating a subset of polygenic risk with any phenotype of interest in family data.

PrgmNr 3199 - Comparing Gene Expression Across Paired Human Airway Models for Cystic Fibrosis Precision Medicine

View session detail

Author Block: **G. He**[1,2], N. Panjwani[1], J. Avolio[1], H. Ouyang[1], S. Keshavjee[3,2], J. M. Rommens[1,2], T. Gonska[1,2], T. J. Moraes[1,2], L. J. Strug[1,2]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada, [3]Univ. Hlth.Network, Toronto, ON, Canada
Disclosure Block: **G. He:** None.
Cystic fibrosis (CF) is caused by loss-of-function variants in the *CF transmembrane conductance regulator (CFTR)*, with modifier genes impacting lung disease severity and therapeutic efficacy. Cultured human nasal epithelia (HNE) are becoming an important surrogate airway model for the gold standard cultured human bronchial epithelia (HBE) to assess the efficacy of CF therapies, because HNE are more easily accessible from patients. However, it remains unknown whether the HNE and HBE genome-wide transcriptome are similar, which we investigate here. RNA-sequencing of paired HNE and HBE samples, cultured and fresh (n=71), that were collected from 21 individuals with CF was carried out. We implemented an equivalence testing procedure based on the two one-sided t-test (TOST) to assess the statistical evidence for similarity in transcriptome between HNE and HBE. A comparison of cultured and fresh airway tissues showed that the culturing process had little impact on the expression level of CF lung disease modifier genes identified in genome-wide association studies (FDR

PrgmNr 3200 - Current methods integrating variant functional annotation scores have limited capacity to improve the power of genome-wide association studies

View session detail

Author Block: **J. Gao**[1], O. Espin-Garcia[2,1], A. D. Paterson[3,1], L. Sun[4,1]; [1]Div. of Biostatistics, Dalla Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada, [2]Dept. of Biostatistics, Princess Margaret Cancer Ctr., Toronto, ON, Canada, [3]Hosp for Sick Children, Toronto, ON, Canada, [4]Univ. of Toronto, Toronto, ON, Canada

Disclosure Block: **J. Gao:** None.

Functional annotations have the potential to increase the power of genome-wide association studies (GWAS) by prioritizing variants according to their biological function. Focusing on convenient variant-specific meta-functional scores including CADD (Kircher et al. 2014) and Eigen (Ionita-laza et al., 2016), we broadly examined GWAS summary statistics of 1,132 traits from the UK Biobank (Sudlow et al., 2015) using the weighted p-value approach (Genovese et al., 2006) and stratified false discovery control (sFDR) method (Sun et al., 2006). These 1,132 traits were rated by Benjamin Nealeâ□□s lab from the Broad Institute as having medium to high confidence for their heritability estimates. Averaged across the 1,132 traits, sFDR was more robust to uninformative meta-scores, but the weighted p-value method identified slightly more variants using CADD (or Eigen). We also considered the recent FINDOR method (Kichaev et al., 2019), which leverages a set of 75 individual functional annotations into GWAS. An earlier application of FINDOR to 27 selected traits from the z7 category defined by Nealeâ□□s lab (SNP-heritability testing p-values

PrgmNr 3201 - Determining disease co-occurrence architecture of Hypertensive Heart Disease in Penn Medicine Biobank using longitudinal EHR data linked with PMBB participants

View session detail

Author Block: **P. Singhal**[1], M. A. Guerraty[1], D. Kim[2], D. J. Rader[1], M. D. Ritchie[1], A. Verma[1,3]; [1]Univ. of Pennsylvania, Philadelphia, MA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Corporal Michael Crescenz VA Med. Ctr., Philadelphia, PA

Disclosure Block: **P. Singhal:** None.

An estimated 1.2 billion individuals worldwide suffer from hypertension. Hypertensive heart disease (HHD) refers to structural alterations, such as left ventricular hypertrophy (or left atrial enlargement) in the heart due to prolonged hypertension and can lead to ischemic heart disease, heart failure (HF), and arrhythmias. HHD represents a transitory subclinical condition between the onset of hypertension and overt cardiovascular disease, typically remaining asymptomatic until the time of the adverse cardiac event. This leads to under-diagnosing of HHD and limits the window of clinical intervention to prevent adverse cardiac outcomes. Moreover, the risk of HHD-related diseases, such as HF, is heightened due to hypertension-associated comorbidities such as diabetes, dyslipidemia, and renal failure. The complex etiology underlying HHD phenotypes is poorly understood and thus it is challenging to predict outcomes in patients. The goal of this study is to leverage longitudinal electronic health record (EHR) data to identify the conditions and comorbidities that influence HHD adverse outcomes for prognostication. We defined an HHD patient cohort (n = 4,268) in the Penn Medicine BioBank dataset using diagnoses codes for HHD (ICD-10: I11*). We constructed a disease-disease map of HHD comorbidities using the Ising model, a type of Markov Random Field undirected probabilistic graphical model. We estimated the co-occurrence of disease pairs by calculating pairwise similarities between diseases as relative risk and conducting network analyses. Our results reflect clusters of both known disease-disease connections, including interactions between circulatory, endocrine, and nervous system diseases, as well as novel pairwise disease interactions between genitourinary, digestive, and musculoskeletal diseases. Our findings shed light on the effect of comorbid conditions on the trajectory of HHD adverse outcomes. Using the probabilistic model of disease co-occurrence from these results, we can identify key comorbidities that predict adverse outcomes among HHD patients, allowing for timely clinical intervention. This work lays the foundation to integrate rare and common genetic variants into the Ising model using a gene-based binning approach. Specifically, incorporating rare loss-of-function variants will enable us to evaluate shared genetic connections between HHD comorbidities. In conclusion, we used EHR data to identify the genetic and non-genetic connections between the comorbidities of HHD to develop a predictive metric for the prevention and clinical monitoring of HHD.

PrgmNr 3202 - Genome-wide association study and heritability of Gulf War illness

View session detail

Author Block: **J. Vahey**[1,2], X. Qin[3,2], A. Stone[4], W. C. Carter[4], L. M. Griffin[4], S. Pyarajan[5], G. Turner[6], E. Gifford[1,2], K. Sims[2], C. D. Williams[2], E. R. Hauser[7,2]; [1]Duke Univ., Durham, NC, [2]U.S. Dept. of VA, Durham, NC, [3]Duke Univ, Durham, NC, [4]U.S. Dept. of VA, Little Rock, AR, [5]VA/HMS, Lexington, MA, [6]U.S. Dept. of VA, Boston, MA, [7]Duke Univ Med Ctr, Durham, NC
Disclosure Block: **J. Vahey:** None.
BackgroundThe Gulf War Era Cohort and Biorepository (GWECB) consists of DNA and survey data from individuals who served in the United States Armed Forces in 1990-1991. Many Veterans who were deployed to the Persian Gulf during this era report pervasive chronic symptoms; the constellation of these symptoms has been defined as Gulf War illness (GWI). There are two main research definitions of GWI: the Kansas and CDC definitions, both of which have variants, for a total of four research case definitions. The etiology of GWI is unknown, although there is evidence that some military exposures are risk factors for GWI. There are no biomarkers for GWI and treatments are very limited. Prior genetic studies of GWI have been limited in scope and heritability of GWI is unknown. **Methods**Self-reported symptoms and health conditions captured by the GWECB survey determined GWI case status (CDC, CDC severe, Kansas, Kansas Symptom Criteria, and Kansas exclusionary criteria). Genotyping was performed using the Illumina Omni2.5-8v1.4 microarray. Plink 1.9 was used for genetic quality control and association analysis. Age, sex, and two genetic principal components, accounting for 79% of genetic variability, were used as covariates for the association analysis. Heritability was calculated using GCTA v1.93.2 using the reml command on autosomes only, with covariates age, sex, and genetic principal components 1-10. **Results**The sample size was 1260 after quality control. The four case definitions examined were CDC GWI (82.6% prevalence in GWECB), CDC Severe GWI (21.7%), Kansas GWI (34.6%), and Kansas Symptom Criteria (68.0%). GWAS for all four of these case definitions were completed, with no results that show genome-wide significance. Of these four definitions, CDC Severe GWI was the most heritable, showing 63% heritability in the adjusted model. CDC GWI and Kansas GWI Symptom Criteria showed heritability close to zero, while Kansas GWI was 4%. **Conclusions**We have completed the first genome-wide genetic analysis and heritability calculation for GWI. Estimating heritability demonstrates that while there are several commonly accepted research definitions of GWI, the case definition used has a big impact on estimates of heritability and thus on gene discovery in GWI. The heritability calculations indicate that CDC Severe GWI is heritable and is a good candidate for future genetic work. The GWAS and heritability results indicate a need for a larger sample size for future genetic analyses, which is available through the Million Veteran Program.

PrgmNr 3203 - Genome-wide study on longitudinal MRI outcomes in 3,513 multiple sclerosis cases across six clinical trials highlights a potential role for MS susceptibility loci in progressive biology

View session detail

Author Block: **N. Sadhu**[1], S. J. Loomis[1], E. Fisher[1], A. Gafson[1], E. Hughes[1], A. Herman[2], R. A. Rudick[3], S. John[1], X. Jia[2], T. R. Bhangale[2], P. G. Bronson[1]; [1]Biogen, Cambridge, MA, [2]Genentech, South San Francisco, CA, [3]Optimal Brain Hlth.Consultants, Jupiter, FL
Disclosure Block: **N. Sadhu:** None.
Introduction: Multiple sclerosis (MS) is an immune-mediated disorder of the CNS characterized by relapse biology, insidious progressive injury, and highly variable outcomes. While the genetics of MS risk is well characterized, the genetics of progression is not well understood and could inform patient stratification and drug discovery. Here, we investigate the genetics of quantitative, longitudinal brain MRI outcomes: a) brain volume change (BVC), a measure of progressive tissue loss; and b) T2 lesion volume change (T2LVC), a measure of inflammatory lesion accumulation.
Methods: We performed genome-wide association studies (GWAS) of BVC (N=2,861) and T2LVC (N=3,513) across 6 randomized controlled trials (RCTs) (mean age 40 years; 63% female). Four RCTs included relapsing MS: DECIDE, OPERA I & II, and ADVANCE (T2LVC only). Two RCTs included progressive MS (PMS): ASCEND (secondary PMS), ORATORIO (primary PMS). ASCEND, ADVANCE, DECIDE were genotyped (Thermo Fisher Axiom array) and imputed. ORATORIO, OPERA I & II were whole genome sequenced (Illumina HiSeq). BVC was annualized percent change in whole BV from 6 months to last time point (min 1.4 years). T2LVC was annualized change in total T2LV from first to last time point (min 1.6 years). Mean duration in study was 1.8 years. Analyses were adjusted for age, sex, ancestry, and treatment arms. Summary statistics from each RCT were pooled for meta-analysis, gene-based tests, and comparison with MS risk loci (IMSGC 2019).
Results: Out of 197 MS risk loci, 41 and 54 non-MHC, autosomal, significant SNPs met our inclusion criteria for BVC and T2LVC meta-analyses (4+ RCTs, heterogeneity I^2MERTK (rs57116599, $p=3.6x10^{-3}$), and *TBX6* (rs3809627, $p=3.3x10^{-3}$). Gene-based tests supported association of *TBX6* with BVC ($p=3.1x10^{-4}$) in this region. Genome-wide analysis of T2LVC revealed a novel association in the regulatory region of *NEDD4L* (rs11398377-GC/G, $p=9.3x10^{-8}$), which showed a corresponding gene-based peak ($p=1.5x10^{-5}$).
Discussion: This study represents the largest RCT-based GWAS of longitudinal MS brain MRI outcomes to date. Our results suggest that *MERTK* and *TBX6* may play a role in progressive brain tissue loss. Genome-wide analysis of MS lesion accumulation also identified ubiquitin ligase *NEDD4L*, a regulator of ion channels involved in excitatory synaptic transmission, highlighting a potential link between immune and neurological pathways. Replication in an independent cohort is needed. Nonetheless, these results contribute to a better understanding of genetic factors affecting progression and inflammatory activity in MS.

PrgmNr 3204 - High quality phased whole genome sequence across patient cohorts is achievable and informs genetic understanding of complex traits

View session detail

Author Block: **S. Mastromatteo**[1], A. Chen[1], J. Gong[1], F. Lin[1], B. Thiruvahindrapuram[1], W. W. L. Sung[1], J. Whitney[1], G. Kaur[1], Z. Wang[1], K. Keenan[1], A. Halevy[1], N. Panjwani[1], L. Sun[2], F. Ratjen[1], J. M. Rommens[1], L. J. Strug[1]; [1]The Hosp. for Sick Children, Toronto, ON, Canada, [2]Univ. of Toronto, Toronto, ON, Canada

Disclosure Block: **S. Mastromatteo:** None.

Epidemiological studies that rely on genotyping arrays and short read whole genome sequencing fail to capture the complete diploid nature of the human genome as they provide little to no information regarding the phase relationships of heterozygous allele pairs. Disregard for the *cis*/*trans* relationship between heterozygous variants misses key data that informs phenotype-genotype relationships. One of the most well-known examples comes from compound heterozygosity in monogenic disorders such as Cystic Fibrosis (CF).

Modern sequencing approaches enable phasing from a single individual by using reads that simultaneously inform multiple heterozygous sites to establish the phase relationship between alleles. Long-read sequencing technologies such as PacBio SMRT sequencing and Oxford Nanopore produce reads that span multiple heterozygous variants. Technologies such as 10x Genomics linked-read technology (10XG) use a standard short-read sequencing pipeline with an additional experimental step that introduces long-range information into the read data. We used Genome in a Bottle (GIAB) data to benchmark the phasing capabilities of PacBio, Nanopore and 10XG against the NA12878 Platinum Genome phase calls. 10XG demonstrated the lowest error rate (0.037%) and long phase blocks (N50=4.19 Mb) in addition to being cost-effective on the scale of a cohort study. We also generated a high-quality composite phase call set for the public reference individual HG002 by combining six different technologies made available by GIAB; 99.996% of heterozygous variants in the high-confidence GIAB VCF were phased in 81 blocks with a phase block N50 of 90.3 Mb.

We sequenced 358 individuals from the Canadian CF Gene Modifier Study cohort using 10XG and observed an average phase block N50 of 4.5 Mb and >98% of all genes shorter than 100kb were phased in a single block. The 10XG data was used to study a GWAS-suggestive locus within chr7q35 known to influence meconium ileus risk in CF. The ~66kb region is complicated by a common 20kb deletion polymorphism and contains repeated homologous genes: *PRSS1*, *PRSS2* and three pseudogenes. We used the 10XG reads to call and phase variation against alternative contig KI270803.1, which includes the 20kb of sequence that is absent from the HG38 chr7 main chromosome. We found that the deletion polymorphism was significantly associated with meconium ileus risk and that it was also a significant *PRSS2* eQTL. The colocalization of these two signals suggests a link between meconium ileus and expression of *PRSS2*. 10XG data enabled complete characterization of this complex region and was critical in generating novel insights into this locus.

PrgmNr 3205 - Leveraging Genetics and Genomics Data to Identify Potential Transcriptional Pathways of Body Mass Index: The Framingham Heart Study

View session detail

Author Block: **H. Xu**[1], A. E. Justice[2], K. L. Mohlke[3], A. M. Cawley[4], A. L. DeStefano[1], I. Dinsmore[5], A. Kollu[6], D. Levy[7,8,9], T. Mirshahi[5], C. D. Still[10], L. A. Cupples[1,11], K. E. North[12], C-T. Liu[1]; [1]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [2]Population Hlth.Sci., Geisinger Hlth., Danville, PA, [3]Dept. of Genetics, Univ. of North Carolina, Chapel Hill, NC, [4]Univ. of North Carolina, Chapel Hill, NC, [5]Molecular and Functional Genomics, Geisinger Hlth.System, Danville, PA, [6]Saint Louis Univ. Coll. of Publ. Hlth.and Social Justice, St. Louis, MO, [7]Dept. of Med., Boston Univ. Sch. of Med., Boston, MA, [8]Framingham Heart Study, Framingham, MA, [9]Population Sci. Branch, Natl. Heart, Lung, and Blood Inst., NIH, Framingham, MA, [10]Geisinger Obesity Inst., Geisinger Hlth.System, Danville, PA, [11]NHLBI and Boston Univ. Framingham Heart Study, Framingham, MA, [12]Dept. of Epidemiology, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC

Disclosure Block: **H. Xu:** None.

Background: Genome-wide association analyses have identified numerous body mass index (BMI) related loci. However, most underlying mechanisms from SNP to BMI remain unknown. Leveraging OMICs data through integrative analyses could potentially provide more comprehensive views of biological pathways.

Methods: We performed a correlated meta-analysis using BMI data, genotype data and array-based transcript expression level data from fasting peripheral whole blood samples in up to 5,619 samples of Framingham Heart Study (FHS). We started with 3,992 SNPs that are in LD (r^2 >0.8) with 97 previously reported BMI variants (P-8) and focused our analyses on 20,692 transcripts with a start position within 1 Mb. We performed separate association analyses of transcript with BMI and transcript with SNP, yielding respective association p-values P_BMI and P_SNP. Then we aggregated these two association results by conducting a correlated meta-analysis that corrects the traditional meta-analysis for the inflation of type-I error when encountering correlation, leading to a meta-analyzed p-value (P_meta). We screened the results to identify transcripts that meet five criteria: P_meta-6, P_BMI-6, and at least nominal association between BMI and SNP (P_BMI-SNPResults: We found 172 SNP-transcript-BMI associations. Among these 172 associations, we identified four unique genes (*NT5C2*, *YPEL3*, *ZNF646*, and *C9orf5*) that were potentially involved in transcriptional pathways from SNP to BMI, where *YPEL3* and *ZNF646* were in the same region (16p11.2). 115 variants were involved in the SNP-transcript-BMI associations for *NT5C2*, including the reported BMI variant rs11191560. *NT5C2* encodes a hydrolase that participates in cellular purine metabolism. The other 56 SNP-transcript-BMI associations for *YPEL3* or *ZNF646* included two reported BMI variants rs4787491 and rs9925964. Our previous experimental data from Drosophila melanogaster knockdown studies support *YPEL3* as the functional gene influencing BMI. At the *C9orf5* locus, we have pinpointed the SNP-transcript-BMI association to the specific reported BMI variant rs6477694.

Summary: By performing a correlated meta-analysis using BMI, genotype and transcript expression level data from FHS, we found four genes (*NT5C2*, *YPEL3*, *ZNF646*, and *C9orf5*) located in three separate regions that may play important roles in transcriptional pathways from SNP to BMI. Further investigation with colocalization analysis and validation in various tissues (adiposity, liver and brain) are underway to better understand the underlying biological processes.

PrgmNr 3206 - Major sex differences in allele frequency for X-chromosome variants in the 1000 Genomes phase 3 data

View session detail

Author Block: **A. D. Paterson**[1], Z. Wang[2], L. Sun[3]; [1]Hosp for Sick Children, Toronto, ON, Canada, [2]Fudan Univ., Shanghai, China, [3]Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **A.D. Paterson:** None.
We identified an unexpectedly high proportion of SNPs on the X-chromosome in the 1000 Genomes phase 3 dataset that have significant sex differences in allele frequencies (SDAF). This SDAF persists for some SNPs in the recently released 1000 Genomes high coverage whole genome sequence, and it is consistent between the five superpopulations. Our primary analysis focused on biallelic SNPs in the 1000 Genomes phase 3 dataset that did not overlap with indels and with global MAF â¥5%. We analysed 222,872 SNPs from the non-pseudo-autosomal region (NPR), 13,244 from the pseudo-autosomal region 1 (PAR1), 634 from PAR2, and 9,075 from the controversial PAR3. We obtained p values from testing for SDAF and to be conservative used p

PrgmNr 3207 - Maternal and parent-of-origin gene-environment effects on the etiology of orofacial clefting

View session detail

Author Block: **N. Rasevic**[1], M. Rubini[2], K. Burkett[1], P. A. Mossey[3], B. Peterlin[4], M. F. J. Khan[2], A. Ravaei[2], L. Autelitano[5], M. C. Meazzini[5], J. Little[1], M-H. Roy-Gagnon[1]; [1]Univ. of Ottawa, Ottawa, ON, Canada, [2]Univ. of Ferrara, Ferrara, Italy, [3]Univ. of Dundee, Dundee, United Kingdom, [4]Univ. of Ljubljana, Ljubljana, Slovenia, [5]Univ. of Milan, Milan, Italy
Disclosure Block: **N. Rasevic:** None.
Objective: Our goal was to investigate maternal and parent-of-origin gene-environment interaction effects on the risk of orofacial clefts for candidate SNPs.
Methods: Genome-wide SNP genotypes were obtained for case-parent triads from the EUROCRAN and ITALCLEFT studies. Candidate SNPs were selected from a previous genome-wide association study, along with surrounding SNPs, resulting in a total of 2142 SNPs. Non-genotyped SNPs were imputed. A total of 411 case-parent triads and 25 case-parent dyads were analyzed using log-linear models to test for maternal and parent-of-origin effects, and their interaction with maternal smoking and maternal use of supplements containing folic acid. False discovery rate adjusted q-values were calculated.
Results: An association was detected for a region in the *ATXN3* gene as it pertains to the interaction between maternal periconceptional smoking and maternal genetic effects (minimum q-value in region = 0.025). Some regions displayed nominally significant associations in the interaction between maternal use of folic acid supplements and parent-of-origin genetic effects, specifically those located in the *CSMD1*, *KCNK10* and *SYT17* gene loci.
Conclusion: Investigation of interactions between environmental exposures and maternal genetic and parent-of-origin genetic effects has detected the potential involvement of regions in the *ATXN3*, *CSMD1*, *KCNK10* and *SYT17* genes. A mutation of the *ATXN3* gene has been involved in a neurological disorder, and the latter three genes have high mRNA expression levels in the brain, which suggest that further investigation of the link between orofacial clefts and brain development would be warranted.

PrgmNr 3208 - Mining biopsychosocial attributes in major depressive disorder by a multi-modal latent topic model

View session detail

Author Block: **M. Wang**[1], X. Meng[1,2], Y. Li[1]; [1]McGill Univ., Montreal, QC, Canada, [2]Douglas Res. Ctr., Montreal, QC, Canada
Disclosure Block: **M. Wang:** None.
Aims: Major depressive disorder (MDD) has been widely accepted as a serious public health issue that negatively impacts an individualâ□□s quality of life and significantly contributes to the global burden of disease. The literature has accumulated ample evidence for MDDâ□□s clinical heterogeneity, which limits the ability to accurately diagnose and treat it. Also, the clinical heterogeneity issue is closely related to the complex interactions between genetic and psychosocial environmental factors. Different genetic architectures across different populations increase the difficulty of articulating how genetic and psychosocial factors are combined to predict the risk of MDD. In the present study, we aim to identify meaningful MDD topics fully considering both genetic and psychosocial factors by using a latent topic model. ***Methods***: Data analyzed are from a longitudinal community-based cohort from Southwest Montreal - Zone dâ□□Ã□pidÃ©miologie Psychiatrique du Sud-Ouest de MontrÃ©al (ZEPSOM) study. The sample size included in this present study was 1083 Caucasians. The ZEPSOM study was designed to capture information on mental health outcomes and their relevant psychological and social attributes. We adopted a multi-modal latent topic model (MixEHR) to explore meaningful combinations of psychosocial data across diverse data structures and different data sources. MixEHR is a multi-view Bayesian topic model that projects each individualâ□□s clinical heterogeneity considering his/her high-dimensional psychosocial data. We then used the identified MDD meta-phenotypes to explore potential genetic variants that were statistically associated with MDD meta-phenotypes. ***Results***: We found that MDD meta-phenotypes based on psychosocial attributes demonstrated superior prediction accuracy compared to the raw psychosocial attributes. We discovered new genetic variants that were not able to be identified by the binary clinical MDD diagnosis. The SNP heritability of lifetime MDD meta-phenotypes was 0.167 (SE=0.305), compared to 0.000001 (SE=0.297) for binary lifetime MDD defined by the structured questionnaire. ***Conclusions***: The present study firstly applied the multi-modal topic model (MixEHR) to infer/identify latent clinical subpopulations/clinical heterogeneity of MDD and discover meaningful genetic and psychosocial attributes that are closely related to its clinical manifestations and etiology. The findings of the present study direct the personalized medicine by providing marginal risk for an individual with those identified biopsychosocial attributes of MDD.

PrgmNr 3209 - Patterns of consanguinity and associations between autozygosity and clinical phenotypes in British South Asians

View session detail

Author Block: **D. Malawsky**[1], E. S. van Walree[2], E. Arciero[1], K. A. Hunt[3], S. Dogra[4], N. Small[5], J. Wright[4], D. Mason[4], R. C. Trembath[6], C. Griffiths[7], D. Posthuma[2], S. Finer[7], D. van Heel[3], H. C. Martin[1]; [1]Wellcome Trust Sanger Inst., Hinxton, United Kingdom, [2]Vrije Univ.it Amsterdam, Amsterdam, Netherlands, [3]Blizard Inst., Queen Mary Univ. of London, London, United Kingdom, [4]Bradford Teaching Hosp. NHS Fndn. Trust, Bradford, United Kingdom, [5]Univ. of Bradford, Bradford, United Kingdom, [6]King's Coll. London, London, United Kingdom, [7]Queen Mary Univ. London, London, United Kingdom
Disclosure Block: **D. Malawsky:** None.
Consanguinity results in high levels of autozygosity, which have been shown to associate with clinically-relevant phenotypes (Clark et al., 2019) as well as rare recessive disorders. We investigated consanguinity patterns and associations between autozygosity and common diseases in >30,000 British Pakistani and Bangladeshi people from two cohorts: Born in Bradford and Genes & Health (G&H). We developed an algorithm that infers an individualâ□□s degree of consanguinity based on their distribution of runs of homozygosity (ROHs). 58% of Pakistanis and 18% of Bangladeshis were inferred to be offspring of second cousins or closer. Consanguinity is less common in younger than in older British Bangladeshi people ($p=4x10^{-6}$), but the opposite is true for British Pakistani people ($p=1x10^{-5}$), suggesting changing marriage practices across time. Amongst British Pakistani people in Bradford, we observe that different kinship networks (biraderis) have differential consanguinity practices. In some groups nearly all individuals are inferred to have related parents, while in others it is 67%. Using electronic health records from G&H, we tested associations between ICD-10 codes and the fraction of the genome in ROHs (F_{ROH}), controlling for population structure and social confounders. Importantly, we included parental relatedness inferred by our method (a proxy for consanguinity-related social factors) as a covariate. We permuted individuals with high identity-by-descent values to control for the effects of fine-scale population structure. After multiple testing correction, we find that F_{ROH} is significantly associated with overall health, as measured by the number of subchapters of the ICD-10 tree with at least one code in an individual (relative risk=1.15 for F_{ROH} equivalent to first-cousin offspring; p=0.0002). Higher F_{ROH} also positively correlated with disorders of the respiratory (RR=1.37; p=0.0004), digestive (RR=1.27; p=0.0009), and endocrine (RR=1.20; p=0.0005) systems. These findings were replicated in UK Biobank with p-9) and a ROH at 3p21.31 associated with chronic rheumatic heart disease (OR=1.21; $p=5.37x10^{-8}$). Overall, we demonstrate multiple novel associations between autozygosity and disease phenotypes, while better controlling for social confounders and fine-scale population structure than previous studies.

PrgmNr 3210 - Prioritizing risk factors of heart failure from UK Biobank using explainable artificial intelligence

View session detail

Author Block: **A. Fungtammasan**, C-F. Lin, P. Nguyen, D. Quang, Y-C. Hwang, C-S. Chin; DNAnexus, Mountain View, CA
Disclosure Block: **A. Fungtammasan:** None.
Understanding and predicting diseases using genetic variation, phenotypic traits, or environmental factors is one of the ultimate goals in human genetic studies. With the recent advancement in machine learning techniques and the availability of large-scale genotype-phenotype databases, sophisticated machine learning methods have become widely adopted by the scientific community for predicting diseases. Many of these techniques offer more flexibility in terms of model assumption compared to traditional statistical methods and the ability to utilize a large amount of data as they have been developed to handle a larger amount of data in other scientific domains. However, the lack of interpretability and explainability of these models faces challenges in detecting errors and building trust.
We demonstrate the use case of explainable artificial intelligence in predicting heart failure, one of the global major causes of death. We experimented with the state-of-art black-box modeling tools such as LightGBM and XGBoost and estimated the contribution of each predictive feature using SHAP. In addition, we also used a recently developed glass-box machine learning model, Explainable Boosting Machine, which directly measures the impact of each predictor and their interactions instead of indirect inference. We found that the ranking of the most contributing risk factors changes with model type, data filtering, and missing value imputation approaches. Nonetheless, at least half of the models reported the higher heart failure rate in males, old age, sedentary lifestyle, and high BMI, which were previously reported in the literature, among the top five risk factors.

PrgmNr 3211 - Stability of polygenic scores across discovery genome-wide association studies

View session detail

Author Block: **L. M. Schultz**[1], A. Merikangas[1,2], K. Ruparel[2], S. Jacquemont[3], D. C. Glahn[4,5], R. E. Gur[2,1], R. Barzilay[2,1], L. Almasy[1,2]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]CHU Sainte justine Res. Ctr., Montreal, QC, Canada, [4]Boston Children's Hosp., Boston, MA, [5]Harvard Med. Sch., Boston, MA

Disclosure Block: **L.M. Schultz:** None.

Polygenic scores (PGS) are commonly evaluated in terms of their predictive accuracy at the population level by the proportion of phenotypic variance they explain. To be useful for precision medicine applications, they also need to be evaluated at the individual patient level when phenotypes are not necessarily already known. Most genomic research has focused on European ancestry individuals, so it is also important to examine the behavior of PGS in other ancestry groups. Hence, we investigated the stability of PGS by using PRS-CS to compute PGS separately for both European (EUR)- and African (AFR)-ancestry individuals in the Philadelphia Neurodevelopmental Cohort (PNC) and the Adolescent Brain Cognitive Development (ABCD) cohort from multiple EUR- and AFR-ancestry discovery meta-GWAS for post-traumatic stress disorder (PTSD), type-2 diabetes (T2D), and height. We then compared PGS in the same individual from one discovery GWAS to another. For the EUR-ancestry discovery GWAS, all pairs of GWAS for the same trait had genetic correlations > 0.92, and SNP-chip heritability on the observed scale ranged from 0.017 Â± 0.003 (PTSD) to 0.697 Â± 0.067 (height). PGS stability was higher for GWAS that explained more of the trait variance, with height PGS being more stable across discovery GWAS than PTSD or T2D PGS. The correlations between PGS computed from two same-ancestry GWAS were lower than anticipated given the strong genetic correlations, ranging from r = 0.378 for PTSD PGS for the ABCD EUR-ancestry cohort to r = 0.736 for height PGS for the PNC EUR-ancestry cohort. Correlations between PGS computed using two different-ancestry GWAS ranged from r = 0.000867 for PTSD PGS for the EUR-ancestry ABCD cohort to r = 0.404 for height PGS for the EUR-ancestry ABCD cohort. Focusing on the upper end of the PGS distribution, different discovery GWAS did not consistently identify the same individuals in the upper quantiles, with the best case being 60% of individuals above the 80th percentile of PGS overlapping from one EUR height GWAS to another and a worst case of fewer than 5% of the same individuals identified as being above the 95th percentile of PGS calculated using different-ancestry GWAS for T2D or PTSD. The degree of overlap decreases sharply as higher quantiles, less heritable traits, and cross-ancestry GWAS are considered. PGS computed from different discovery GWAS have only modest correlation at the level of the individual patient, underscoring the need to proceed cautiously with integrating PGS into precision medicine applications.

PrgmNr 3212 - Systematic Comparisons for Composition Profiles Taxonomic Levels and Machine Learning Methods for Microbiome based Disease Prediction

View session detail

Author Block: **K. Song**[1], Y-H. Zhou[2], F. A. Wright[3]; [1]North Carolina State Univeristy, Raleigh, NC, [2]Cary, NC, [3]North Carolina State Univ, Raleigh, NC
Disclosure Block: **K. Song:** None.
Microbiome composition profiles generated from 16S rRNA sequencing have been extensively studied for their usefulness in phenotype trait prediction, including for complex diseases such as diabetes and obesity. These microbiome compositions have typically been quantified in the form of Operational Taxonomic Unit (OTU) count matrices. However, alternate approaches such as Amplicon Sequence Variants (ASV) have been used, as well as the direct use of k-mer sequence counts. The overall effect of these different types of predictors when used in concert with various machine learning methods has been difficult to assess, due to varied combinations described in the literature. Here we provide an in-depth investigation of more than 1,000 combinations of these three clustering/counting methods, in combination with varied choices for normalization and filtering, grouping at various taxonomic levels, and the use of more than ten commonly used machine learning methods for phenotype prediction. The use of short k-mers, which have computational advantages and conceptual simplicity, is shown to be effective as a source for microbiome-based prediction. Among machine-learning approaches, tree-based methods show consistent, though modest, advantages in prediction accuracy. We describe the various advantages and disadvantages of combinations in analysis approaches, and provide general observations to serve as a useful guide for future trait-prediction explorations using microbiome data.

PrgmNr 3213 - Two large-scale genome sequencing studies show a consistent association between mitochondrial DNA copy number and personality traits

View session detail

Author Block: **R. Oppong**[1], Y. Qian[1], T. Tanaka[1], A. Z. Moore[1], A. Terracciano[2], F. Cucca[3], M. Picard[4], D. Schlessinger[1], L. Ferrucci[1], J. Ding[1]; [1]Natl. Inst. on Aging, Baltimore, MD, [2]Dept. of Geriatrics, Florida State Univ., Tallahassee, FL, [3]IRGB CNR, monserrato, Italy, [4]Dept. of Neurology, Columbia Univ., New York, NY

Disclosure Block: **R. Oppong:** None.

Levels of mitochondrial DNA copy number (mtDNAcn) in tissues and blood can be altered in conditions of stress and illness such as diabetes, major depression, and peripheral artery disease, suggesting that additional synthesis of mtDNA can occur completely separate from mitochondrial biogenesis. And in the situation of stress, mtDNA can be released from mitochondria into cells and eventually spill into the circulation, indicating that mtDNAcn in the blood may correlate with mtDNA released under stress. Since specific personality traits make individuals more or less susceptible to environmental stress and thus prone to developing pathology and premature death, we hypothesize that mtDNAcn is a biomarker of stress burden and could mediate the association between personality and excess mortality. To test this hypothesis, we studied the relationship between personality (assessed by the *big five* personality domains and facets of the Revised NEO Personality Inventory), levels of mtDNAcn estimated by whole genome sequencing, and mortality in participants of the Baltimore Longitudinal Study of Aging (BLSA; 800 ids with complete data, mean age 75 ranging from 48 to 100, 48% women). Replication of results was sought in the SardiNIA Project (SardiNIA; 600 ids with complete data, mean age 57 ranging from 15 to 96, 62% women). Our results show that mtDNAcn is negatively associated with the Neuroticism domain and positively associated with the Extraversion domain and facets within Agreeableness and Conscientiousness domains. We demonstrate the robustness of these associations by showing that the direction and size of the effects are robust to adjustment for potential confounders such as age, sex, platelet count, and white blood cell parameters. In accordance with results from the literature, we show that participants with depressive symptoms, assessed using the Center for Epidemiologic Studies Depression Scale (CES-D), had lower mtDNAcn than those with no depressive symptoms. Finally, we report an effect of personality on mortality and demonstrate that this effect is fully mediated via mtDNAcn. These results confirm our hypothesis that mtDNAcn is a biomarker in the biological pathway that connects certain personality characteristics with mortality. To our knowledge, this is the first study that shows a clear association between mtDNAcn and personality, and it establishes a new connection between mitochondrial outcomes and psychosocial stressors and emotional states that is highly relevant to human health.

PrgmNr 3214 - Unexpected somatic mosaicism and reduced penetrance: *NF1* pathogenic variants are found in 1 in 850 individuals

View session detail

Author Block: **T. Drivas**[1,2,2], T. Nomakuchi[1], M. Bonanni[3], A. Raper[2], S. M. Kallish[4,2], K. L. Nathanson[5], M. D. Ritchie[3]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]The Univ. of Pennsylvania, Philadelphia, PA, [3]Univ. of Pennsylvania, Philadelphia, PA, [4]Children Hosp Philadelphia, Philadelphia, PA, [5]Univ Pennsylvania SOM, Philadelphia, PA

Disclosure Block: **T. Drivas:** None.

Neurofibromatosis type 1 (NF1), caused by pathogenic variants in the *NF1* gene, is a relatively common autosomal dominant RASopathy thought to affect roughly 1 in 3,000 individuals. Although variable expressivity in NF1 is well-documented, the condition is thought to exhibit near-complete penetrance. We have recently been referred a number of patients with incidentally discovered pathogenic variants in *NF1* with minimal features of the syndrome on exam or history. This led us to hypothesize that the true incidence of pathogenic variants in the *NF1* gene might be higher than reported, with reduced penetrance, a broader phenotypic spectrum, a higher incidence of somatic mosaicism/clonal hematopoiesis of indeterminate potential than has previously been reported, or some combination thereof. We investigated this hypothesis in the Penn Medicine Biobank, which contains electronic health record and exome sequencing data on 43,731 patients of the Penn health system. We identified 51 individuals with clearly pathogenic *NF1* variants, equating to an incidence of 1 in 850, more than 3 times higher than commonly accepted estimates. Only 19 of these 51 individuals had an *NF1* diagnosis in their medical record, suggesting that two thirds of these individuals may be undiagnosed. Comparing variant allele frequencies (VAF) of the *NF1* variants between NF1-diagnosed and undiagnosed patients, we found that patients with a known NF1 diagnosis had an average VAF of 0.47, consistent with germline variation, while the average VAF in the undiagnosed group was only 0.28, implying that in at least some subjects in this group the *NF1* pathogenic variant might exist in a somatic mosaic state. We next conducted a phenome-wide association study (PheWAS) to identify phenotypes more common in the 51 patients with *NF1* pathogenic variants in an unbiased way. Numerous significant associations were identified, but interestingly these associations were entirely driven by the patients with a known NF1 diagnosis, suggesting that the presence of an *NF1* pathogenic variant in blood does not, *per se*, increase risk for disease. This is in line with our clinical experience, where we have seen that patients with incidentally discovered *NF1* pathogenic variants often do not exhibit features consistent with an NF1 diagnosis. In summary, an unbiased investigation of the population-level incidence and penetrance of *NF1* pathogenic variants suggests that these variants are far more common than reported, with a high degree of somatic mosaicism, and significantly reduced penetrance. The further investigation of this finding will be critical for accurate counseling of patients and families.

PrgmNr 3215 - Variants in nicotinic acetylcholine receptor and dopaminergic genes in relation to smoking relapse and proportion of time in relapse throughout adulthood in female nurses

View session detail

Author Block: **S. Jones**[1], B. J. Wolf[1], A. J. Alberg[2]; [1]Med. Univ. of South Carolina, Charleston, SC, [2]Univ. of South Carolina, Columbia, SC
Disclosure Block: **S. Jones:** None.
Background: Smoking cessation and behaviors are heritable. Single nucleotide polymorphisms (SNPs) of nicotinic acetylcholine receptor (nAChR) and dopaminergic genes have been linked to short-term cessation or cross-sectional smoking status, but studies have been limited by short-term follow-up (â¤ 90 days) which cannot account for multiple quit attempts and relapse events over time, a common smoker phenotype. **Methods:** Based on prior short-term studies, ten SNPs in nAChR or dopaminergic genes were selected, six in *CHRNA5-A3-B4*, one in *CHRNB2*, two in *DRD2*, and one in *COMT,* to test associations with odds of smoking relapse and proportion of time in relapse in adulthood in women of European ancestry. The study population was from the Nursesâ Health Study (NHS) and NHS-2, two longitudinal cohorts of female nurses. Participants included 12,060 ever smokers who reported having quit smoking at â¥1 timepoint with existing quality-controlled genotype data. Follow-up time was counted after the first report of smoking cessation with a median follow-up of ~32 years. A zero-inflated beta regression was used to model SNP associations with odds of relapse during follow-up and, conditional on relapse, model SNP associations with proportion of time in relapse. **Results:** Women with AA genotypes compared to those with AG or GG genotypes for *CHRNA5* SNP rs16969968 G>A or *CHRNA3* SNP rs1051730 G>A had 16% increase odds of smoking relapse in adulthood (p = 0.04 for both), but neither SNP was associated with the proportion of time in relapse. Further, among women who relapsed, four other SNPs were associated with the proportion of time in relapse; these included more time in relapse for women with CC genotypes (versus CT or TT genotypes) of *CHRNA5* SNPs rs588765 T>C (p = 0.04) and rs680244 T>C (p = 0.05) and AA genotype (versus AG or GG genotypes) for *DRD2* SNP rs6277 G>A (p = 0.01), whereas time in relapse was less for those with AA or AG genotypes (versus GG genotype) for *COMT* SNP rs4680 G>A (p = 0.03). **Conclusion:** Prior genetic association studies of smoking relapse have been limited by short-term follow-up. This study revealed that some SNPs were associated with relapse over decades of follow-up, whereas other SNPs were not associated with odds of relapse but were associated with the amount of time in smoking relapse during adulthood. The findings emphasize the value of long-term follow-up to characterize smoking relapse phenotypes as this approach can identify smokers with high-risk genotypes for long-term relapse prevention interventions. This may have important translational implications as smoking-related health risk increases with duration of smoking.

PrgmNr 3216 - *NUDT15* Variants Associated with Thiopurine Toxicity in 1,643 Pediatric Leukemia Patients

View session detail

Author Block: **A. M. Muir**[1], Z. E. Fan[1], K. Cao[1], F. Lin[1], M. Luo[1,2], Y. Zhong[1,2], L. F. Surrey[1,2], G. B. Wertheim[1,2], J. Schubert[1], J. Wu[1], E. A. Fanning[1], J. Chen[1], E. H. Denenberg[1], S. R. Rheingold[3,2], S. P. MacFarland[3,2], S. K. Tasian[3,2], S. P. Hunger[3,2], M. M. Li[1,2,3]; [1]Dept. of Pathology and Lab. Med., Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [3]Div. of Oncology and Ctr. for Childhood Cancer Res., Children□□s Hosp. of Philadelphia, Philadelphia, PA

Disclosure Block: **A.M. Muir:** None.

Thiopurines are commonly used anti-cancer drugs in the treatment of pediatric acute lymphoblastic leukemia (ALL) and lymphoma (LL). Nudix Hydrolase 15 deficiency due to germline genetic polymorphisms in *NUDT15* gene is associated with impaired thiopurine chemotherapy metabolism and potential toxicity. Risk polymorphisms in *NUDT15* occur with a collective minor allele frequency of ~4.5% in the general population and >10% in East Asians. Preemptive screening for known risk alleles allows for genotype-guided dosing of thiopurine chemotherapy, an effective strategy for mitigating severe myelosuppression and hepatotoxicity. In this study, we report *NUDT15* sequencing results in a cohort of 1,643 patients with *de novo* or relapsed hematologic malignancies (HMs) whose specimens underwent targeted *NUDT15* sequencing (n=750) or HM gene panel (n=893) at the Children□□s Hospital at Philadelphia (CHOP). The self-declared racial and ethnic ancestry of our cohort broadly reflects the racial diversity of the United States: 52.8% Caucasian (non-Hispanic), 12.6% Black (non-Hispanic), 13.4% Hispanic/Latino (any race), and 6.1% Asian. Targeted *NUDT15* analysis was performed by Sanger sequencing of all coding exons including intron-exon boundaries. The next-generation sequencing based HM gene panel analyzed 119 HM-associated genes, including *NUDT15*. Known *NUTD15* risk alleles were identified in 8.5% of patients (139/1,643). The higher frequency of our cohort compared to the general population is likely due to â□□detection biasâ□□ amongst non-CHOP specimens, many of which were sent for targeted pharmacogenomic testing given clinical signs of thiopurine intolerance in the patients. The *NUDT15* *2 ([c.50_55dupGAGTCG(;)c.415C>T]; 41.0%) and *3 (c.415C>T, 30.2%) were the most common risk alleles in our cohort with *2 allele highly prevalent in patients of Hispanic ethnicity,*3 and *6(c.50_55dupGAGTCG) alleles in Caucasians, and *2, *3, and *6 in Asians. Ten patients (0.6%) had biallelic variants in *NUDT15*; patients with such genotypes generally manifest with extreme intolerance to thiopurine drugs. In addition, eight rare *NUDT15* missense variants and one in-frame deletion (c.460_468delAACCATCTG) were identified. At least one of these patients required 6-MP dose reduction due to hematologic toxicity. Our results identifying *NUDT15* risk alleles in a large cohort of children with HMs highlight the need for routine *NUDT15* pharmacogenetic testing prior to thiopurine-based chemotherapy administration. Such an approach is essential for optimizing individualized thiopurine dosing to mitigate undesirable hematopoietic and hepatic toxicity during ALL/LL therapy.

PrgmNr 3217 - A case report of rare t(8;22)(p11.2;q13) positive AML

View session detail

Author Block: **N. Owen**[1,2], W. Bi[1,2], J. L. Smith[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics, Houston, TX
Disclosure Block: **N. Owen:** None.
Acute myeloid leukemia (AML) is a genetically heterogeneous form of hematologic cancer with hundreds of known cytogenetic abnormalities driving malignancy. One rearrangement, t(8;22)(p11.2;q13), is a rare finding initially described over 20 years ago. It is a variant of the much more common t(8;16)(p11.2;p13) that creates a fusion protein between KAT6A (MOZ), a histone lysine acetyltransferase and CREBBP, a transcription factor, the effect of which is transcriptional dysregulation. This rearrangement is most common as a therapy-related finding in patients that have been treated with chemotherapy previously for solid tumors or other hematological diseases. The t(8;22) is thought to create a fusion between KAT6A and EP300, a protein with high homology to CREBBP. It is a much rarer phenomenon, so much remains unknown about this abnormality. Interestingly, while the t(8;16) shows a female preponderance (~2/3 of patients), all five cases of t(8;22) to date in the literature have been male. More patients with this translocation are necessary to fully understand the clinical picture and prognosis associated with this translocation, as well as the similarities and differences between it and related variant translocations.
Here we present a case of a 67-year- old male who presented with AML. Initial cytogenetic workup identified t(8;22)(p11.2;q13) as the sole anomaly in 16/20 metaphases analyzed. Five weeks later, a bone marrow sample was clear of cytogenetically abnormal cells. Seven weeks after that, a third sample was once again positive for t(8;22) with significant clonal progression. Four clones were observed, the first containing the t(8;22) and trisomy 8. Trisomy 8 is a common numerical abnormality in myeloid malignancies and is the most common secondary finding in t(8;16) positive AML. The second clone contained the abnormalities in the stemline, and also contained trisomy 5 and an isochromosome of 8p replacing one chromosome 8; this was the most prevalent clone in the sample. The third clone contained only the derivative (22) from the t(8;22), and had structural abnormalities of 8p and 10q. The fourth clone contained all of the abnormalities in clone 3, and had additionally acquired a reciprocal translocation between 2p and 5q. The patient passed away soon after. The prognosis associated with t(8;16) has been described as poor, though newer literature suggests the presence of de novo AML and non-complex karyotypes improve that prognosis. With so few cases described to date, the prognosis associated with t(8;22) has not been clearly established. This case suggests that t(8;22) may share the same prognosis as its counterpart t(8;16).

PrgmNr 3218 - A case study of AML in a patient with an additional *RUNX1T1* signal and a loss of one 3' *CBFB* signal in 88% of the nuclei examined

View session detail

Author Block: **F. Guirales**[1,2], A. Okabe[2,3], B. Hamid[2,3], A. Wang[2,3], J. King[4], M. Guardiola[4], K. Eastwood[4], C. A. Tirado[4,2]; [1]William Paterson Univ., Wayne, NJ, [2]The Intl. Circle of Genetics Studies, Los Angeles, CA, [3]Univ. of California Los Angeles, Los Angeles, CA, [4]Baylor Scott and White Hlth., Dept. of Pathology, Temple, TX

Disclosure Block: **F. Guirales:** None.

Acute myeloid leukemia (AML) is a heterogeneous blood cancer characterized by the proliferation of immature myeloid cells. Herein we present an 18-year old female with marked leukocytosis (WBC of ~300 x 10^9/L), thrombocytopenia with spontaneous intracerebral and cerebellar bleeds and anemia. Mother is noted to have *CHEK 2* mutation. Half sister died of AML S/P stem cell transplant. The bone marrow aspirate smears show numerous blasts, atypical monocytes and rare eosinophilic precursors with large basophilic granules. The blasts display highly irregular nuclei and dispersed chromatin and prominent nucleoli. The bone marrow biopsy is hypercellular (90-100%) and effaced by sheets of immature cells with irregular nuclei and moderate cytoplasm. The flow cytometry reveals the blast population expresses monocytic markers CD14, dim CD4, HLADR, CD13, CD33, CD15, CD36, CD64, CD38, CD11b with a subset co-expressing CD34 and CD117 and aberrantly expressing CD7 and CD5. This population is negative for CD16 and CD56. Chromosome analysis reveals an abnormal karyotype with a derivative trisomy 8 and a derivative chromosome 16 involving an inversion 16q as well as deletion 16q in 21 of the 22 metaphase cells examined. Thus, the karyotype was described as 47,XX,+der(8)add(8)(q24.3),der(16)inv(16)(p13.1q22)del(16)(q22)[21]/ 46,XX[1]. DNA FISH analysis revealed abnormalities for *RUNX1T1* (8q21.3) and *CBFB* (16q22) genes. There was a loss of one 3' *CBFB* signal in 88% of the nuclei examined. There was also an additional *RUNX1T1* signal in 83.5% of the nuclei examined. This was confirmed also in the concurrent chromosome study where there was an extra copy of chromosome 8. These findings correlate with concurrent conventional cytogenetic findings and were described as nuc ish(RUNX1T1x3,RUNX1x2)[167/200],(5'CBFB'x2,3'CBFBx1)(5' con 3'x1)[176/200],(EVI1,TAS2R1,EGR1,DEK,RELN,TES,ABL1,NUP214,KMT2A,PML,RARA,PTPRT,MYBL2,BCR) x2[200]. The National Comprehensive Cancer Network (NCCN) guidelines for AML state that rearrangements of the CBFB gene place the patient in the favorable risk category. An extra copy of chromosome 8 is usually seen in AML with rearrangements of 16q. Clinical correlation was suggested.

PrgmNr 3219 - A rare inv(14)(q11.2q32.3) in Acute Lymphoblastic Leukemia

View session detail

Author Block: **X. Zhao**[1,2], N. Owen[1,2], J. Kubala[2], W. Bi[1,2], J. L. Smith[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics, Houston, TX
Disclosure Block: **X. Zhao:** None.
Introduction: The inv(14)(q11.2q32.3) has been studied extensively in TCRÎ±/Î´-TCL11 fusion positive T-cell Prolymphocytic Leukemia (T-PLL). However, only a few cases have been documented with a cytogenetically identical translocation in patients with Acute Lymphoblastic Leukemia (ALL). These cases are thought to contain a distinct CEBPE-IGH fusion. Due to the scarcity of cases, the disease spectrum and clinical significance of inv(14)(q11.2q32.3) in ALL is still largely unknown. **Methods:** Retrospective review of the cytogenetic and FISH results on a consecutive cohort of 11,120 patients studied at Baylor Genetics from 2010 to 2021 was conducted. The cohort included 5314 females and 5806 males, with a concern for hematologic malignancy. Cytogenetic analysis of bone marrow or leukemic blood samples was performed using standard methods, and nomenclatures of chromosome abnormalities were based on ISCN. The involvement of IGH in the patient identified in this study was confirmed using a commercially available break-apart probe. The probe for CEBPE (RP11-147E17) was grown as a BAC according to standard protocol of the laboratory. **Results:** The patient is a 22-year-old female with a diagnosis of ALL. Her initial cytogenetic workup by FISH and chromosomes was normal. Fifteen months later, a small subclone (8.5%) was identified with an IGH rearrangement detected only by FISH. Five months after that, FISH was 46% positive for an IGH rearrangement and an abnormal clone (11/20 cells) with inv(14)(q11.2q32.3) was identified in the chromosome analysis. A reciprocal translocation between 1p and 6q was also seen. Further FISH studies identified the rearrangement partner for IGH to be CEBPE. One month later the abnormal clone was undetectable after treatment, but the patient relapsed 3 months after that with the same clone; no new clonal abnormalities were observed. **Discussion:** Only a few cases of inv(14) or the analogous t(14;14) have been published. The long term outcomes of patients with this inversion remain to be explored. To our knowledge, this is the first case of a female pediatric ALL driven by an inv(14) with an CEBPE-IGH rearrangement; previously all reported pediatric cases were male. The identification of this one patient in our large cohort has reiterated the rarity of this abnormality and allowed us to track the clinical course of a patient with this inversion over time. In addition, caution should be paid to distinguish these from T-PLL patients when interpreting the results. Further investigation is warranted to understand the disease mechanisms.

PrgmNr 3220 - Does it make sense to report single pathogenic alleles in cancer susceptibility genes for childhood cancer patients? Results from the BASIC3 and Texas KidsCanSeq studies

View session detail

Author Block: **S. E. Plon**[1], L. R. Desrosiers[1], S. Scollon[2], A. K. Petersen[3], H. DAI[4], J. Reuther[5], G. Miles[6], D. M. Muzny[4], A. Roy[5], S. Kulkarni[7], R. A. Gibbs[4], A. L. McGuire[8], G. E. Tomlinson[9], J. Bernini[1], J. B. Gill[10], T. Griffin[11], K. Vallance[12]; [1]Dept. of Pediatrics, Baylor Coll. Med., Houston, TX, [2]Baylor Coll. of Med., Houston, TX, [3]Randall's Children Hosp. and Legacy Hlth., Portland, OR, [4]Baylor Coll. Med., Houston, TX, [5]Pathology, Baylor Coll. Med., Houston, TX, [6]Molecular and Human Genetics, Baylor Coll. Med., Houston, TX, [7]Baylor Genetics, Houston, TX, [8]Baylor Col Med, Houston, TX, [9]UT Hlth.San Antonio, San Antonio, TX, [10]UT MD Anderson Cancer Ctr., Houston, TX, [11]Children's Hosp. of San Antonio, San Antonio, TX, [12]Cook Children's Med. Ctr., Fort Worth, TX

Disclosure Block: **S.E. Plon:** Non-remunerative positions or influence such as officer, board member, trustee, or public spokesperson; Baylor Genetics.

Background: Multiple studies report that 8-13% of pediatric cancer patients have pathogenic or likely pathogenic variants (PV/LPV) in cancer susceptibility genes, although standards vary for reporting single PV/LPVs in genes associated with autosomal recessive (AR) syndromes. We report here results from two studies of pediatric cancer patients for the frequency of single AR variant results and relationship to the patientâ□□s presentation. **Methods:** The BASIC3 study included clinical germline and tumor exome sequencing in an ethnically diverse cohort of sequentially diagnosed children with CNS and non-CNS solid tumors. The Texas KidsCanSeq study included germline exome and panel sequencing for pediatric solid tumor patients across six sites in Texas. In both studies, other medical problems noted in the medical record were provided to the reporting laboratory. All result categories on the clinical exome and panel (when available) were returned to parents. **Results:** Among 278 enrolled patients in BASIC3 there was one patient with two PVs (a homozygous variant in *TJP2*) and 18 (6.5%) patients had a single PV/LPV in a wide spectrum of AR cancer genes. Only one patient had a tumor type (Wilms tumor) previously associated with the gene (*DIS3L2*). Patent foramen ovale in a patient with *FANCA* allele was the only syndromic feature. In the first 366 patients in KidsCanSeq there were no patients with biallelic PV/LPVs, 15 patients (4.1%) with one PV/LPV in AR genes of which 2 had the associated tumor type (*DIS3L2* with Wilms tumor and *SLC26A4* with thyroid cancer). Not included here are patients with PV/LPV in genes with both AR and dominant cancer phenotypes e.g., *ATM.* Germline results disclosure with parents discussed (1) their oncologist might request further targeted testing for a 2nd variant if there is concern that the patient has this rare disorder, (2) the patient (and often parent) are carriers of this rare disorder, (3) the lack of evidence relating the variant to the childâ□□s cancer diagnosis and (4) lack of any additional surveillance recommendations.**Conclusions:** There is increasing use of exome, genome or large hereditary cancer panels for genomic evaluation of cancer patients. Recessive diagnoses are very uncommon but about 5% of pediatric cancer patients have a single PV/LPV in rare AR cancer genes without other findings. These test results require substantial efforts for disclosure and family understanding. We suggest that limiting reporting to biallelic patients or those with features of the disorder including the specific cancer type would streamline the process at the cost of reproductive information. BASIC3 and Texas KidsCanSeq were supported by 1U01HG006485.

PrgmNr 3221 - Heterogeneity in cytological presentation of gene amplification and concurrent amplification of *MYC* and *PVT1* in therapy related acute myeloid leukemia (AML)

View session detail

Author Block: **R. Koduru**[1], K. Wilson[1], R. Garcia[2], O. Weinberg[1], P. Patel[1], K. Kirthi[3], W. Chen[1]; [1]Univ Texas Southwestern Med Ctr, Dallas, TX, [2]UT Southwestern, Euless, TX, [3]Med. City Hosp., Dallas, TX
Disclosure Block: **R. Koduru:** None.
The chromosomal region at 8q24.1 is a hot spot for somatic alterations in hematological and non-hematological cancers. Amplification of this region has been reported in about 1% of myeloid malignancies including AML and myelodysplastic syndrome (MDS). Here we present a case of therapy related AML in which *MYC* and *PVT1* were concurrently amplified in an amplicon at 8q24.1 which presented as double minute (dmin), ring chromosome or homogeneously staining region (hsr) in leukemia cells in the same bone marrow (BM) specimen. A 70-year female with a history of breast carcinoma, non-small cell lung cancer, and renal carcinoma, developed therapy related AML. Karyotype at the time of diagnosis was abnormal with 45,X,-X,3-12dmin[cp3]. FISH and molecular studies showed loss of *ABL1* in 27.3% cells, but negative for all other FISH probes for AML, and negative for *FLT3*, *IDH1*/*IDH2*, or *KIT* mutations. Repeated karyotype analysis of BM detected four clones: 46,XX,del(9)(q13q34)[5]/47,idem,+1-2r[9]/45,X,-X,dmin[3]/45,X,-X,HSR(8)(q22)[3]. FISH with *MYC* break apart probe detected amplification on the ring chromosome, the hsr and the dmin. Hybridization with *PVT1* probe also detected amplification on all these abnormal chromosomes. Immunohistochemical study on BM core biopsy showed MYC expression in leukemic blasts. Thus, these results together indicate co-amplification of *MYC* and *PVT1* and expression of *MYC* in leukemia cells. Dmin or hsr constitutes amplification of genomic DNA. In myeloid proliferations frequent site of amplification is 8q24.1 containing *MYC* and less frequently 11q23 containing *KMT2A*. *MYC* amplification is frequently reported in non-therapy related AML, and rarely in therapy related AML. Although this region contained 15 other genes, only alterations, a *NSMCE2*/*PVT1* fusion and *PVT1* amplification were reported. This is the second report of amplification of *PVT1* along with *MYC*. Contrary to the previous findings that amplification was mainly in the form of dmin or ring chromosomes, our case showed plasticity in cytological presentation of the amplicon in the form of dmin, ring chromosome, and hsr, and a relatively simple but abnormal karyotype. The presence of independent clones, one with ring chromosomes, the other with hsr, and a third with only dmin in the same specimen suggests independent genetic events that may have occurred simultaneously in leukemia cells. Alternatively, each of these clones may suggest that hematopoietic stem cells were subjected to different genotoxic events when the patient received treatment for prior malignancies. Consistent with previous findings, the patient ultimately had poor clinical outcome.

PrgmNr 3222 - Imputing missing genotypes and examining rare genetic variants in families with high risk of lung cancer

View session detail

Author Block: **Z. Zhang**; Baylor Coll. of Med., Houston, TX
Disclosure Block: **Z. Zhang:** None.
Lung cancer (LC) is the leading cause of cancer death in USA. Although low dose spiral CT screening can detect early-stage LC patients, even early diagnosed patients still have about a 30% cause-specific mortality. Some individuals with genetic susceptibility express a high risk for LC development, particularly if exposed to tobacco smoke or other carcinogens. However, more complete characterization of genetic risk is necessary before it can be used to stratify individuals by the LC risk and diagnose LC at an early stage. Recently, rare variants with large effects on LC risk were identified. However, due to low allele frequency, rare variants are easier to detect in multiplex families than in sporadic samples, because rare variants, especially the causal variants, tend to accumulate in families. Therefore, data from families having multiple relatives affected with LC is a good option for discovery of rare variants. In this study, we investigated the genetic risk of LC by analyzing families from a unique data resource established by Genetic Epidemiology of Lung Cancer Consortium (GELCC). GELCC has been collecting samples and data from individuals with a strong history of LC for the last 20 years. We have cancer phenotypes, detailed smoking exposure data and biological specimens available from 868 probands, as well as next generation whole exome sequencing (WES) of high-risk LC families. Our studies have identified variants associated with familial lung cancer, and further studies to evaluate the quality of the results and validate findings are ongoing. In addition to identifying risk variants, we also developed an efficient pipeline for family-based imputation and association analysis. Missing genotype data is common in genetic studies of lung cancer with short survival time, especially in family-based studies. Genetic imputation is an efficient tool that is widely used to infer the missing data and boost the power of the analysis. Although population-based imputation is very popular and has very high accuracy for variants with high allele frequency, accuracy declines with allele frequency. Family-based imputation may overcome the drawback; however, it is not as mature as population-based imputation. Many problems are unsolved, including software efficiency and quality control of the post-imputed genotypes, etc. We have used the framework of Morgan 2 and GIGI to fill in missing genotypes using array-based analysis on 249 individuals and linkage information from sparser arrays for an additional 2,432 participants. The association analysis is performed by incorporating the metric of the imputation quality as weights to decrease the false discovery rates.

PrgmNr 3223 - Serial cytogenetic studies showing evolution of trisomy 8 positive CMML to AML with a complex karyotype characterized by deletion 7q and a jumping 1q translocation without evidence of trisomy 8

View session detail

Author Block: **J. L. Smith**[1,2], B. Yuan[3,4], L. R. Smith[5], W. Bi[1,2], N. Owen[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics Lab., Houston, TX, [3]Seattle Children's Hosp., Seattle, WA, [4]Univ. of Washington, Seattle, WA, [5]High Mark Genetics, LLC, Los Alamos, NM
Disclosure Block: **J.L. Smith:** None.
Approximately 30% of Chronic Myelomonocytic Leukemia (CMML) patients have one or more clonal cytogenetic abnormalities. Both trisomy 8 and deletion 7q are common findings in CMML, and both have been classified as intermediate to high risk. Jumping translocations are seen on occasion in cytogenetic studies of hematologic samples, most commonly in multiple myeloma and myeloid malignancies. Most reported cases of myeloid malignancies involve an initial diagnosis of MDS with transformation to AML or worsening MDS. We report a rare patient with a jumping 1q translocation that emerges and disappears during a series of studies as the patientâ□□s disease course evolved from CMML to AML.

The first bone marrow sample from a 48-year-old male had an indication of CML. Chromosome analysis was negative for the Philadelphia chromosome but positive for trisomy 8 in 6/20 cells. FISH testing did not include probes to assay for a BCR/ABL1 fusion, so a cryptic rearrangement could not be excluded; however, FISH with a centromeric probe showed 46% of interphase cells with three copies of the centromere 8. A year later, a second bone marrow sample was submitted, this time with an indication of CMML. Again, the sole abnormality was trisomy 8 [8/20 G-banded cells; 84/200 interphase cells by FISH].

Within 5 months, a third sample with an indication of CMML not having achieved remission was submitted. Neither FISH nor chromosome analysis showed any evidence of trisomy 8; however, 16/20 G-banded cells had a deletion of 7q22q36. Four additional bone marrow samples were submitted 3, 4, 5 and 7 months later before the patient succumbed to his illness. At the time of the fourth sample, the diagnosis was classified as CMML-2 with increasing blasts; three different jumping 1q translocations were seen in different clones, all of which contained a deletion 7q. The three recipient chromosome regions were 14p, 21p and 22p; the most common recipients of the 1q are the short arms of the acrocentric chromosomes. The last three samples were submitted with an indication of AML. Two studies showed the deleted 7q as well as the jumping 1q translocations. Finally, while the fifth sample showed no evidence of the 1q translocations, all 20 G-banded cells and 90% of interphase cells were positive for the deletion 7q.

This case report not only highlights a rare case of CMML with a jumping 1q but affords the opportunity to observe the emergence and disappearance of the 1q translocations over time. Jumping 1q chromosomes are thought to be associated with disease progression and are rarely noted to have disappeared, particularly if the patient remains in relapse without normalization of the karyotype.

PrgmNr 3224 - Somatic Mutation Profiling of Basal Cell Carcinoma in a Population Exposed to Arsenic

View session detail

Author Block: **M. G. Kibriya**[1], F. Jasmine[1], T. Islam[2], A. Ahmed[2], M. Rakibuz-Zaman[2], M. Islam[2], M. Kamal[3], C. R. Shea[1], M. Argos[4], H. Ahsan[1]; [1]Univ Chicago, Chicago, IL, [2]Univ Chicago Bangladesh, Dhaka, Bangladesh, [3]Bangabandhu Sheikh Mujib Med. Univ., Dhaka, Bangladesh, [4]Univ. of Illinois Chicago, Chicago, IL

Disclosure Block: **M.G. Kibriya:** None.

Background: Non-melanoma skin cancers (NMSC) are the most prevalent malignancy in the United States, with an estimated 2 million new diagnoses each year. NMSC includes basal cell carcinoma (BCC) and squamous cell carcinoma. While ultraviolet (UV) radiation exposure and skin sensitivity are known risk factors for NMSC in Caucasians, arsenic (As) exposure may be a major risk factor in other populations. **Aim:** To identify cancer-associated somatic mutations in BCC**,** we conducted next-generation sequencing of DNA samples from BCC (n=26), corresponding blood samples (n=26), and the biopsy samples from normal skin tissue (n=16) from sixteen independent subjects. **Material and Methods:** We utilized tissue samples from individuals participating in the **B**angladesh vitamin **E** and **S**elenium **T**rial (BEST). From BEST participants who developed NMSC over the follow-up period of 8 years and had skin biopsy tissue preserved in RNA Later available, we selected the first 26 BCC cases. Healthy skin tissues (surrounding non-cancerous skin lesions) were taken from an independent set of participants (who had arsenical keratosis) for comparison. All participants were exposed to As-through drinking As-contaminated ground water and had visible As-related skin lesions at the time of trial enrollment. Targeted DNA sequencing of a comprehensive cancer panel of 406 cancer related genes was performed on Illumina platform. **Result:** For each sample, ~14 million clusters were sequenced with ~94% reads with quality score >Q30 (average quality score of 35.9) and post-QC average depth of coverage ~240x. To detect somatic mutation in BCC or healthy skin tissues, we excluded the variants found in the corresponding blood DNA. As expected, we detected a number of somatic mutation in healthy skin tissue. Then to detect the BCC associated somatic mutations, we excluded the variants found in healthy skin. Missense somatic mutations (n=214) were detected in 133 genes. On average, each BCC sample revealed 8 missense somatic mutations. The most common missense mutations were seen in *PRKDC, ABL1, TET1 TAF1L, SYNE1, EP400, SETD2, KMT2C, MTOR* etc. Mutations in genes seen in UV-induced BCC (*PTCH1, SMO, NOTCH1, NOTCH2, TP53* etc.) were also found, but were less frequent in this set of As-associated BCC. **Conclusion:** We report the somatic mutational profile using a large comprehensive panel of targeted genes in As-related BCC. The pattern indicates that a different molecular genomic pathway may be involved in the pathogenesis of As-related BCC. Further studies are needed to confirm the findings.

PrgmNr 3225 - The g-quadruplex stabilizing agent GQC-05 inhibits LINE-1 retroelement activity in human bronchial epithelial cells

View session detail

Author Block: **K. Ramos**[1], P. Bojang, Jr.[2], V. Gokhale[3], L. Hurley[3]; [1]TAMHSC-Inst. of BioSci.s and Technology, Houston, TX, [2]Univ. of Kentucky, Lexington, KY, [3]Univ. of Arizona, Tucson, AZ
Disclosure Block: **K. Ramos:** None.
Nearly 100 copies of the Long interspersed element-1 (LINE-1) retrotransposon remain active in the human genome. These elements function as insertion mutagens and epigenetic effectors in various types of cancer. In silico analyses of LINE-1 sequences identified putative guanine-quadruplex (G-quadruplex) structures within the 3â□□ untranslated region (3â□□ UTR). G-quadruplexes are classified into intramolecular and intermolecular structures based on the number of nucleotides involved in forming the G-quadruplex, with both DNA and RNA able to form these structures. G-quadruplexes are secondary structural motifs formed from the stacking of two or more guanine tetrads into square co-planar structures held together by Hoogsteen type hydrogen bonds of guanines that create a central channel where a cation such as K+ resides to stabilize the structure and neutralize the negative central charge. The present studies were conducted to determine if G-quadruplexes within the 3â□□ UTR participate in the regulation of LINE-1 activity. We found that GQC-05 binds to the G-quadruplex sequence in the 3â□□ end of LINE-1, as evidenced by stabilization at one equivalent by 13Â°C. Drug binding reduced LINE-1 mRNA levels in the absence of any appreciable effect on DNA polymerase activity below a concentration of 50Î¼M and inhibited LINE-1 retrotransposition and clonal expansion in soft agar of human bronchial epithelial cells expressing constitutively active LINE-1. The multilevel efficacy of GQC-05 suggests that modulation of G-quadruplexes in the 3â□□ untranslated region of the LINE-1 may be targeted pharmacologically to modulate tumors plagued by aberrant LINE-1 activity.

PrgmNr 3226 - Underutilization of germline testing for prostate cancer patients: Are genetic testing criteria a tool or an obstacle?

View session detail

Author Block: **S. Nielsen**[1], C. Pieczonka[2], M. Gazi[3], S. Heron[4], D. Cahn[5], L. Belkoff[6], A. Berger[7], B. Mazzarella[8], J. Veys[9], D. Morris[10], A. Engelman[11], P. Dato[12], R. Cornell[13], R. Bevan-Thomas[14], D. Wise[15], M. Hardwick[16], K. Hatchell[17], E. D. Esplin[17], R. L. Nussbaum[18], N. Shore[19]; [1]Invitae, Chicago, IL, [2]Associated Med. Professionals, Chicago, IL, [3]Univ. Urology Associates of New Jersey, Hamilton, NJ, [4]Advanced Urology Inst., St. Petersburg, FL, [5]Colorado Urology, Lakewood, CO, [6]Mid Atlantic Urology, Bala Cynwood, PA, [7]Associated Urological Specialists, Chicago Ridge, IL, [8]Urology Austin, Austin, TX, [9]North Georgia Urology, Dalton, GA, [10]Urology Associates PC, Nashville, TN, [11]Cancer Ctr. of South Tampa, Tampa, FL, [12]Genesis Hlth.Partners, San Diego, CA, [13]Urosurgery Houston, Houston, TX, [14]Urology Partners, Arlington, TX, [15]NYU Langone's Perlmutter Cancer Ctr., New York, NY, [16]Targeted Med. Ed., Cupertino, CA, [17]Invitae, San Francisco, CA, [18]San Francisco, CA, [19]Atlantic Urology Clinics, Myrtle Beach, SC

Disclosure Block: **S. Nielsen:** Salary/Employment; Invitae.

Background: Pathogenic/likely pathogenic (P/LP) germline variants occur in ~10-15% of all prostate cancer (PCa) patients. However, complicated workflows and restrictive testing criteria lead to underutilization of genetic testing. A multi-center study was conducted to determine the incidence of P/LP variants in PCa patients who met (in criteria, IC) and did not meet (out of criteria, OOC) the NCCN 2019 PCa germline genetic testing criteria. **Methods:** An IRB-approved, multicenter, prospective registry was initiated with 15 community and academic urologists nationwide. Consecutive, unselected PCa patients ages 18-90 were consented and underwent an 84-gene germline panel test, with clinical information collected via clinician-completed case report forms. Statistical significance was determined by chi-square test without Yatesâ correction. **Results**: Of the 764 patients to date, 429 (56%) were IC and 335 (44%) were OOC. Excluding carriers of heterozygous variants in genes associated with autosomal recessive cancer conditions (e.g. *MUYTH* heterozygotes), **7.3**% (56/764) of patients had a P/LP variant on an 84-gene panel: **8.2%** (35/429) of IC patients and **6.3%** (21/335) of OOC patients (p=0.11). For a 12-gene PCa panel, P/LP yield was 5.1% (39/764), with no significant difference between IC and OOC patients (p=0.33). Of 56 pts with P/LP variants, 79% (n = 44) were potentially eligible for precision therapy (e.g. PARP inhibitors), clinical treatment trials, and/or other clinical management interventions, 36% (n = 16) of whom wouldâve been missed by criteria. **Conclusions**: There was no statistically significant difference in the yield of P/LP variants between IC and OOC patients, reinforcing that guidelines miss a significant number of patients with P/LP variants. Current criteria for germline testing of PCa patients are an obstacle to identifying patients with precision therapy-eligible P/LP variants and should be revised.

PrgmNr 3227 - Age-specific transcriptional risk scores (TRS) link GWAS to eQTLs and predict therapeutic response across 8 common cancer types

View session detail

Author Block: **E. Hall**[1], P. Awadalla[2]; [1]Ontario Inst. for Cancer Res., Toronto, ON, Canada, [2]Ontario Inst Cancer Res., Toronto, ON, Canada
Disclosure Block: **E. Hall:** None.
Early-onset cancers (age at pathologic diagnosis â¤ 45) represent a distinct biological age subgroup associated with poor prognosis and recurrence risk. We aimed to identify differences in gene expression and their prognostic significance for early-onset primary tumors across eight common cancer types: breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), pancreatic adenocarcinoma (PAAD), prostate adenocarcinoma (PRAD), ovarian serous cystadenocarcinoma (OV), and acute myeloid leukemia (AML). We integrated TCGA and GTEx RNA-seq data to perform differential gene expression (DE) analyses and identify age-related differentially expressed genes (age DE genes) significantly up or down-regulated in early- vs late-onset primary tumors compared to normal tissue for each tumor/tissue type. In parallel, we also identified significant eGenes for each tumor/tissue type by integrating summary-level data from GWAS with data from expression quantitative trait locus (eQTL) studies to identify genes whose expression levels were associated with hereditary cancer risk. We hypothesized that the summation of risk-allele associated gene expression, i.e., a transcriptional risk score (TRS), would provide a superior estimation of cancer risk and prognosis compared to biomarkers identified through differential gene expression analyses. Thus, utilizing either age DE genes or eGenes, we constructed TRS for early-onset cancer risk for each tumor/tissue type. We show that TRS based on eGenes were not only significantly elevated in early- vs late-onset untreated primary tumors, but also served to identify patients who were at high risk of developing disease-specific complications. This relationship between the TRS and risk of disease-specific complications was validated in multiple external cohorts. For example, for BRCA, we identified a 55 gene TRS which was not only significantly elevated in early-onset, early-stage BRCA primary tumors regardless of tumor subtype, but also significantly predicted which patients were at high risk of developing a recurrence following confirmation of disease-free status (p

PrgmNr 3228 - Broadly pleiotropic effects of pathogenic variants in tumor suppressor *CHEK2*

View session detail

Author Block: **L. D. Ward**[1], M. M. Parker[1], A. Deaton[2], H-C. Tu[1], A. Flynn-Carroll[2], R. Hoffing[2], C. Willis[3], G. Hinkle[2], P. Nioi[1]; [1]Alnylam, Cambridge, MA, [2]Alnylam Pharmaceuticals, Cambridge, MA, [3]Alnylam Pharmaceuticals, Boston, MA

Disclosure Block: **L.D. Ward:** Salary/Employment; Alnylam Pharmaceuticals.

Exome sequencing has been performed on most UK Biobank participants, allowing genome-wide and phenome-wide association studies of rare coding variants. In an exome-wide burden test of association between rare predicted damaging variants and menopause timing in the UK Biobank, we found associations with five genes (*CHEK2*, *RAD54L*, *DCLRE1A*, *HROB*, and *TOP3A*), all of which were involved in the DNA-damage repair pathway. Of these genes, the strongest association was between rare predicted damaging variants in *CHEK2* and later menopause (1.46 years; p = 6.2 Ã 10^{-51}). *CHEK2* is a moderate-penetrance breast cancer gene and is known to have pleiotropic associations with other cancers (in the case of Li-Fraumeni syndrome), but the association with later menopause was unexpected. Through causal mediation analysis, we estimate that approximately 8% of the breast cancer risk of *CHEK2* is conferred through delayed menopause. We then performed a phenome-wide association study of rare predicted damaging variants in *CHEK2* to test for additional associated phenotypes in 363,977 men and women in the UK Biobank. We found that in addition to breast cancer (OR = 1.87; p = 6.2 Ã 10^{-51}) and prostate cancer (OR = 1.67; p = 2.8 Ã 10^{-12}), *CHEK2* associates with increased platelet crit (beta = 0.18 SD; p = 5.0 Ã 10^{-39}) and leukocyte count (beta = 0.16 SD; p = 4.6 Ã 10^{-29}), and increased risk of several noncancer diseases: uterine fibroids (OR = 1.49; p = 4.3 Ã 10^{-10}), polycystic ovarian syndrome (OR = 3.05; p = 1.5 Ã 10^{-8}), prostate hyperplasia (OR = 1.40; p = 2.5 Ã 10^{-8}), and seborrheic keratosis (OR = 1.31; p = 6.4 Ã 10^{-7}). While the hematological associations have been described previously, we additionally show in a stratified analysis that they are found in both cancer-free women and men. These results suggest a function for *CHEK2* in protecting against not only cancer, but also other syndromes involving hyperproliferation of tissue, consistent with its role in the cell cycle.

PrgmNr 3229 - Genome-wide interaction analysis identified low-frequency variants with sex disparity in lung cancer risk

View session detail

Author Block: **Y. Li**[1], X. Xiao[1], J. li[1], J. Byun[1], C. Cheng[1], R. Hung[2], C. Amos[1], INTEGRAL-ILCCO consortium; [1]Baylor Coll. of Med., Houston, TX, [2]Lunenfeld-Tanenbaum, Toronto, ON, Canada
Disclosure Block: **Y. Li:** None.
Introduction: Lung cancer is the leading cause of cancer death for both men and women worldwide with a complex genetic and molecular mechanism. Differences by sex in lung cancer incidence and mortality have been reported. There is growing evidence showing that sex disparity in lung cancer risk cannot be fully explained by sex differences in smoking behavior, implying existence of genetic and molecular basis for sex disparity in lung cancer development. However, the information about sex dimorphism in lung cancer risk is quite limited despite the great success in lung cancer association studies.**Methods:** By adopting a stringent two-stage analysis strategy, we performed a genome-wide gene-sex interaction analysis using genotype from a lung cancer cohort including ~ 47,000 individuals with Caucasian ancestry. The significant variants were further submitted to allelic-specific expression using genotype and gene expression data in lung tissue from 244 men and 133 women in GTEx. Two public functional annotation tools, CADD and RegulomeDB were applied for functional inference of the candidate variants.**Results:** Three low-frequency variants (minor allele frequency Conclusions: We conducted a large-scale gene-sex interaction scanning in lung cancer and we identified three significant variants with different risk effects based on sex. Our study is one of the first studies to examine sex disparity in lung cancer development and our results provided insights about the genetic and molecular mechanism underlying the differences in lung cancer susceptibility between men and women. Our findings also suggested quite a few variants still remain undetected in lung cancer, including those with low allele frequency requiring a very large sample size; or the variants affecting disease risk thorough interaction with environmental and biological factors requiring more robust analysis strategy.

PrgmNr 3230 - GWAS of Uveal Melanoma Reveals Novel Genome Wide Significant Locus

View session detail

Author Block: **N. Tsao**[1], H. Thomsen[2], K. G. Ewens[3], S. Coupland[4], K. Hemminki[5], S. M. Damrauer[6], A. GANGULY[7]; [1]Univ. of Pennsylvannia, Philadelphia, PA, [2]Heidelberg Univ., Heidelberg, Germany, [3]Univ Pennsylvania Sch Med, Philadelphia, PA, [4]Univ. of Liverpool, Liverpool, United Kingdom, [5]Univ. of Heidelberg, Heidelberg, Germany, [6]Hosp. of the Univ. of Pennsylvania, Philadelphia, PA, [7]Univ. of Pennsylvania, Philadelphia, PA

Disclosure Block: **N. Tsao:** None.

Background: Uveal melanoma (UM) is a rare intraocular tumor. It is also the most common form of adult-onset primary ocular malignancy with a significantly high degree of mortality associated with distant metastasis. Most cases of UM are sporadic with only a very small fraction presenting with a family history of the disease. The only gene known to be associated with UM is *BAP1*. We performed a genome-wide association study (GWAS) to identify common genetic variants associated with US. **Methods:** We conducted a case control analysis of individuals with tissue confirmed UM and age and sex matched controls that were free of cancer diagnoses based on ICD-9/10 diagnosis codes in their electronic health record, which we identified from the Penn Medicine Biobank. All individuals were genotyped using the Axiom Precision Medicine Research Array (PMRA) and the data imputed to the HRC reference panel using the University of Michigan imputation server. We tested the association of 4,146,536 DNA sequence variants for an association with UVM using logistic regression controlling for age, sex, and population structure using PLINK 2.0. **Results:** There were 399 individuals with UM and 748 without an EHR diagnosis of cancer. Individuals with UM had a mean age of 57.2 and 215 (53.%) were male. The mean age of those without cancer was 54.9, and 48.4(%) were male. We replicated the previous association on chromosome 5p rs421284(OR = 1.39, p = 1.93E-4). Our analysis also identified a single, novel genome wide significant locus on chromosome 5q, rs6862372 (OR = 0.2, p = 7.9 E-11). rs6862372 is a significant cis-eQTL for Pelota MRNA Surveillance And Ribosome Rescue Factor (PELO) (P=1.1E-10) and Integrin Subunit Alpha 1 (IGTA) (P=9.1E-7) based on data from eQTLgen. IGTA1 has previously been shown to be upregulated in metastatic uveal melanoma cells. **Conclusion:** In conclusion, our GWAS identified a novel locus that implicates PELO and IGTA1 could play a role in uveal melanoma oncogenesis. We aim to replicate these findings using local and international cohorts and further these analyses using a time to event frame work to investigate the genetics of uveal melanoma metastasis progression.

PrgmNr 3231 - Predicted gene expression identifies novel genes associated with lung cancer in African Americans

View session detail

Author Block: **M. Betti**[1], J. Jaworski[1], J. S. Rao[2], B. Ryan[3], A. G. Schwartz[4], J. K. Wiencke[5], M. A. Bruce[6], S. Zhao[1], S. J. Chanock[3], M. C. Aldrich[1], J. Hellwege[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Univ. of Miami, Miami, FL, [3]Natl. Cancer Inst., Bethesda, MD, [4]4Karmanos Cancer Inst., Wayne State Univ., Detroit, MI, [5]Univ. of California, San Francisco, San Francisco, CA, [6]Univ. of Mississippi Med. Ctr., Jackson, MS

Disclosure Block: **M. Betti:** None.

African Americans have an elevated risk of developing lung cancer compared to Whites, despite lower average cigarette consumption. Recent research has begun to address this paradox with studies attempting to specify genetic mechanisms contributing to this racial disparity with genome-wide association study (GWAS) of lung cancer risk in African Americans. We extended this work with two key innovations, the use of TOPMed imputed single nucleotide polymorphism (SNP) data, as well as the assessment of genetically-predicted gene expression. We performed a GWAS using 1,612 cases and 3,294 controls from the African American Lung Cancer cohort. Genotype data was imputed using the TOPMed Imputation Server, expanding the number of variants from 932,188 to 15,415,642. Associations between SNPs and lung cancer diagnosis were evaluated using a logistic regression model, adjusted for sex, age, smoking status, study site, and proportion of Yoruban ancestry, as determined via ADMIXTURE with 1000 Genomes CEU and YRI reference data. The four SNPs reaching genome-wide significance were near *CHRNA5* within the 15q25.1 locus (lead SNP rs17486278, $p = 4.46 \times 10^{-10}$, odds ratio = 1.30 [95% Confidence Interval: 1.19-1.42]). We observed suggestive (p-6) associations at seven other loci.

We next aimed to link GWAS hits with tissue-specific changes in gene expression using GTEx eQTL data. To boost statistical power, we implemented a joint-tissue imputation (JTI) approach to leverage information across the transcriptomes of different GTEx tissues based on shared chromatin regulatory profiles. Combining GWAS summary results with S-PrediXcan and JTI models resulted in significant gene-tissue pairs (p -7) for 6 genes: *SNRNP200* (significant across 17 tissues), *CEP295* (significant across 6 tissues), *CSKMT* (significant across 4 tissues), *FCSK* (significant across 2 tissues), *CCL28* (significant in tibial artery), and *RPF1* (significant in skeletal muscle). *SNRNP200* was significant in lung tissue ($p = 1.72 \times 10^{-12}$, effect size = 0.703).

We confirmed significant associations between lung cancer risk and GWAS-identified SNPs and discovered tissue-specific predicted gene expression using eQTL data. This work expands our knowledge of the regulatory mechanisms underlying these GWAS signals and contributes to our understanding of the genetic architecture of lung cancer in African Americans. This represents a step forward towards addressing the historic underrepresentation of African ancestry populations in genetics research.

PrgmNr 3232 - SNPs at SMG7 associated with time from biochemical recurrence to prostate cancer death

View session detail

Author Block: **R. J. Klein**[1], X. Song[1], M. Ru[1], Z. Steinsnyder[2], K. Tkachuk[2], Z. GÃ¼mÃ¼Å[1], K. Offit[2], V. Joseph[2]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Mem. Sloan Kettering Cancer Ctr., New York, NY

Disclosure Block: **R.J. Klein:** None.

Previously, in a genome-wide association study we identified several loci with genetic variants associated with prostate cancer survival time. To validate these loci and examine whether they influence the entire course or only a particular stage of the disease, we genotyped two SNPs in de-identified DNA samples of 1298 men who underwent radical prostatectomy at Memorial Sloan Kettering Cancer Center (MSKCC) from January 1990 through January 2006. We tested SNPs rs2702185 and rs73055188 for association with prostate-cancer-specific survival time using a Cox proportional hazard model, adjusted for age, PSA level, Gleason score and stage at surgery and stratified by Ashkenazi Jewish (AJ) or non-Jewish, European ancestry. We further tested the SNP rs2702185 for association with time to biochemical recurrence and biochemical recurrence to death from disease in the non-AJ population. We performed linkage disequilibrium (LD) analysis to identify SNPs in high LD with rs2702185, and examined their functional effects, chromatin accessibilities and gene expressions.

Here, we report an intergenic variant in SMG7, rs2702185, was associated with prostate-cancer-specific survival time in the non-AJ patient population. Decomposing the survival time into two intervals delineated by biochemical recurrence, we identified rs2702185 specifically associated with time from biochemical recurrence to prostate cancer death in the non-AJ patient population. Nine variants were in LD with rs2702185. Of these, rs10737246 was most likely functional, based on differential LD with rs2702185 between AJ and northwest European populations and overlap with open chromatin. Notably, both the tissue distribution of open chromatin around rs10737246 and eQTL analysis using the eQTL Catalogue points to a role for this SNP in immune (primarily T cells) rather than prostate cells. The potential causal effect of rs10737246 is yet to be experimentally validated.

In conclusion, we found that rs2702185 in SMG7 is associated with time from biochemical recurrence to prostate cancer death, and its LD partner rs10737246 is predicted to be functional.

PrgmNr 3233 - Cisplatin induced gene expression across three different cancer cell line types reveal common and distinct molecular regulatory effects

View session detail

Author Block: **M. Luck**[1], C. Del Greco[2], A. L. Stark[3]; [1]Univ. of Notre Dame, Notre Dame, IN, [2]Univ. of Michigan, Ann Arbor, MI, [3]Univ. of Notre Dame, South Bend, IN
Disclosure Block: **M. Luck:** None.
Cisplatin chemotherapy treatment, although a widely utilized and often initially successful anti-cancer treatment, is challenged with immediate or acquired resistance and harmful side effects that can persist once treatment is stopped. This study aims to characterize the molecular effects of cisplatin treatment in breast, lung, and colorectal cancer cell lines and to investigate the induced role of SNAIL regulated genes following cisplatin treatment across different cancer types. Ten genes of interest regulated by SNAIL were selected based on high-fold expression changes following Snail up-regulation, and qPCR was used to assess expression changes 24- and 48- hours following cisplatin exposure in MCF7 (breast), A549 (lung), and HCT116 (colorectal) cancer cell lines. Across the 10 genes (S100P, DECR1, THBS1, TACSTD2, H2AFJ, TMEM47, HOXA9, UPK2, TSPYL5, and AKT3), we observed different patterns of induced expression ranging from consistently induced significant (p

PrgmNr 3234 - Computational evidence of BRCA1-NC3 receptor interaction and druggability analysis in BRCA1-related oncogenesis

View session detail

Author Block: **M. Herceg**[1], M. L. Brown[2], A. Shehu[3]; [1]George Mason Univ. Sch. of Systems Biology, Fairfax, VA, [2]George Mason Univ. Ctr. for Drug Discovery, Fairfax, VA, [3]George Mason Univ. Dept. of Computer Sci., Fairfax, VA
Disclosure Block: **M. Herceg:** None.
Breast cancer susceptibility gene 1 (*BRCA1*) on human chromosome 17q21 was identified in 1994 as the site of breast cancer-causing mutations (Ma *et al*. 2005. Oncogene). The canonical pathway by which mutations in *BRCA1* lead to cancer is through impaired DNA repair (reviewed by Tarsounas and Sung. 2020. Nat. Rev. Mol. Cell Biol.). However, this does not explain the hormone-relatedness of *BRCA1* oncogenesis (Rosen *et al*. 2003. Trends Endocrinol. Metab.). Rosen and colleagues characterized a putative Estrogen Receptor-Î± (ER-Î±; *ESR1*)-BRCA1 binding site (2005. Oncogene.). If the ER-Î±-BRCA1 interaction plays a significant role in breast cancer oncogenesis, it could represent a druggable pathway in a subset of breast cancers. By some estimates, 10-36% of *BRCA1*-related breast cancers retain ER function (ER+) (Tung *et al*. 2010. BCR). ER-Î± is a member of the NC3 Nuclear Receptor family. The members of NC3, including ER-Î² (*ESR2*), the Androgen Receptor (*AR*), and the Progesterone Receptor (*PGR*), are steroid receptors with a characteristic fold. There is evidence that BRCA1/2 interacts with at least some other members of the NC3 family, and with AR in particular (Yeh *et al*. 2000. PNAS.). In addition, Rosen, Brown, and colleagues described the disruption of the ER-Î±-BRCA1 complex by drug A7 and several other drugs (2014. Mol. Endocrinol.). They hypothesized that the disruption was due to a conformational change, since the characterized drugs were not found to bind competitively with estrogen. In this work, using informatics techniques, we present evidence of sequence and structural conservation of binding sites identified by Rosen, Brown, and colleagues as important in drug-ER-Î± and ER-Î±-BRCA1 binding across members of the NC3 nuclear receptor family (2014. Mol. Endocrinol.). In addition, using protein-ligand docking simulations, we present evidence on possible cross-binding of A7 and related drugs to various NC3 receptors. Finally, we validate our simulation results with wet-laboratory measurements. Our work represents a step towards understanding the molecular basis of NC3 receptor family positive *BRCA1*-related cancers by shedding light on the possible hormone-dependence mechanism of these cancers. This work also lays the foundation for drug screening and design for these cancers.

PrgmNr 3235 - Identifying functional melanoma risk variants and susceptibility genes via massively parallel reporter assays and multi-QTL in normal melanocytes and malignant melanomas

View session detail

Author Block: **E. Long**[1], J. Yin[1], K. Funderburk[1], J. Feng[1], M. Xu[1], T. Zhang[1], T. Myers[1], A. Golden[1], L. Jessop[1], E. Kim[2], M. M. Iles[3], M. T. Landi[1], M. H. Law[4], S. J. Chanock[1], K. M. Brown[5], J. Choi[1]; [1]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Bethesda, MD, [2]Dept. of Internal Med., Yonsei Univ. Coll. of Med., Seoul, Korea, Republic of, [3]Leeds Inst. for Data Analytics, Sch. of Med., Univ. of Leeds, Leeds, United Kingdom, [4]Statistical Genetics, QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [5]Natl. Cancer Inst., Bethesda, MD

Disclosure Block: **E. Long:** None.

Melanoma is a deadly skin cancer that originates from melanocytes. The largest genome-wide association study (GWAS) so far has identified 54 melanoma-associated loci, but molecular mechanisms for most of them have not been characterized. To identify susceptibility genes/variants from 54 melanoma GWAS loci, we integrated massively parallel reporter assays (MPRA) with expression quantitative trait loci (eQTL) and methylation QTL (meQTL) in normal melanocytes and malignant melanomas. For MPRA, we included 2102 variants based on linkage disequilibrium (LD, R^2 > 0.8) or GWAS statistics (log likelihood ratio 0.1) compared to non-significant variants (P = 0.001). A substantial proportion of MPRA-significant variants (121/285, 42.5%) were predicted to have effects on at least one transcription factor binding site and displayed a correlation between transcriptional activity and predicted allelic binding scores (R^2 = 0.05, P = 0.01). Moreover, a majority of the MPRA-significant variants (172/285, 60.4%) overlapped genome-wide significant eQTLs/meQTLs in melanocytes or melanomas, nominating 31 candidate susceptibility genes. In conclusion, MPRA along with multi-QTL analysis in relevant tumor and normal cell types can quickly prioritize plausible functional candidates from thousands of high-LD variants from melanoma GWAS loci.

PrgmNr 3236 - A deep learning approach for efficiently incorporating genomic data into ovarian cancer prognosis prediction

View session detail

Author Block: **L. Jiang**[1], C. Xu[2], Y. Bai[3], Y. Gong[1], A. Liu[1], Y-P. Wang[3], H-W. Deng[4]; [1]Tulane Univ. Div. of BioMed. Informatics and Genomics, New Orleans, LA, [2]The Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK, [3]Tulane Univ. Dept. of BioMed. Engineering, New Orleans, LA, [4]Tulane Univ., New Orleans, LA

Disclosure Block: **L. Jiang:** None.

With high-throughput sequencing technologies, high-dimensional genomic data are becoming more commonly available. Although combining genomic and clinical information can help us develop a better understanding of complex mechanisms behind cancer patientsâ survival, the issue of low sample size with many current datasets can cause overfitting problem in statistical model fitting and machine learning. Moreover, the association between features of interest (e.g., gene expression) and the survival outcome can be nonlinear but the conventional Cox Proportional Hazard model was not designed to handle nonlinearity. Recently, there has been a rapid growth in the application of deep learning techniques in cancer research. These practices deal with nonlinearity intrinsically, and by virtue of the flexibility in building a deep neural network, structures can be incorporated to extract features from high-dimensional data and prevent overfitting. Here, we propose a two-phase deep learning framework composed of a Beta Variational Autoencoder (Beta-VAE) that encodes information underlying the high-dimensional features into lower-dimensioned latent factors, and a Cox layer that outputs the prognosis index for each patient. In the first phase, Beta-VAE is trained for information extraction and dimension reduction. Beta-VAE is an altered form of Variational Autoencoder (VAE) that allows less information to be passed through the bottleneck layer. In the model we expect the network to only encode the most important information for lowering the reconstruction loss, and potentially further prevent overfitting by removing extra noises. With the trained weights retained, in the second phase, the Cox layer is added and the network is trained for prognosis prediction. By implementing this training procedure, we expect the learned latent factors to be more survival-analysis-associated. We collected data for 451 ovarian cancer patients from The Cancer Genome Atlas (TCGA) database, including their clinical data (e.g., age, ethnic group, clinical stage), gene expression, and miRNA data. We applied our method on single-omics data (e.g., gene expression, miRNA, respectively) to assess which single view of the molecular profile grants more improvement to ovarian cancer prognosis prediction when added on top of clinical data. Furthermore, we also assessed the model performance when integrating multi-omics and clinical data. We compared the performance of this proposed model with other approaches (i.e., Cox-EN, PCA, Cox-PASNet) and found that in general our model yielded the best prediction while requiring relatively short running time.

PrgmNr 3237 - An improved approach leveraging MRI and genomic information for classification and segmentation of glioma lesions

View session detail

Author Block: **H. Suryadevara**[1], X. Wang[2]; [1]Univ. of North Texas, Denton, TX, [2]Univ North Texas, Denton, TX

Disclosure Block: **H. Suryadevara:** None.

Gliomas are one of the most common types of primary brain tumors, which can be broken into low- and high-grade. High grade gliomas are more aggressive as compared to low grade. Gliomas are classified according to the type of glial cell involved in the tumor, as well as the tumor's genetic features. Early detection of gliomas can substantially increase the survival rate of a patient. Magnetic resonance imaging (MRI) is widely used for the detection of brain tumors. However, MRI application is an error-prone and exhaustive activity especially in image recognition. In this study, we introduce a Spatial Pyramid Attention Network (SPANet) to investigate the role of attention blocks for medical image analysis. In contrast to other attention-based networks that leverage global average pooling, the SPANet considers both structural regularization and structural information. The SPA block is flexible to be deployed to various convolutional neural network (CNN) architectures in segmentation and classification. Utilizing SPANet, we propose an improved approach for classification, localization, and segmentation of glioma lesions. The proposed approach for tumor detection is validated on glioma genomic data obtained from The Cancer Genome Atlas (TCGA) and corresponding MRI data from The Cancer Imaging Archive (TCIA). The proposed method achieved greater than 0.85 prediction scores in classification, as well as 0.90 Tversky accuracy rate in segmentation when we use the combined data including T1-weighted MRI scans, FLAIR segmentation masks, and patient genomic biomarkers. Moreover, classification and segmentation outcomes are superior as compared to existing methods.

PrgmNr 3238 - Analyses of breast cancer microbiome data

View session detail

Author Block: **S. Sohail**, M. Burns; Loyola Univ. Chicago, Chicago, IL
Disclosure Block: **S. Sohail:** None.
The microbiome has been implicated as a potential driver in a variety of cancers, including breast cancer [3]. There are numerous studies, with associated microbial sequencing data, available on this topic. [1; 2; 3]. However, microbiome analysis is a rapidly developing field and older analytical methods rapidly become deprecated. In the work presented here we have performed a meta-analysis of the raw data from available studies, with the intended goal of assessing (1) the reproducibility of the original findings and (2) assessing breast cancer-specific microbial signatures that occur across study cohorts [1; 2; 3; 4]. Using the DADA2 pipeline, an amplicon sequence variant approach providing high taxonomic resolution, specificity, and lower false positives than the mixture of now-outdated bioinformatic approaches used in the original studies [5], we have identified major improvements in analysis approaches as well as novel findings that were not revealed in the original reports. We anticipate that these new research results arising from a thorough meta-analysis of existing datasets using modern best-practices and up-to-date databases will provide researchers vital information surrounding the relationship between breast cancer and its associated microbial communities.
[1] Urbaniak et al. 2014. Appl. Environ. Microbiol. 80:3007-3014. [2] Xuan et al. 2014. PLoS One 9:e83744. [3] Hieken et al. 2016. Sci. Rep. 6:30751. [4] Chan et al. 2016. Sci. Rep. 6:28061. [5] Prodan et al. 2020. PLoS One 15:e0227434.

PrgmNr 3239 - Building multi-omics classifier to improve phenotype prediction in heterogeneous cancer data

View session detail

Author Block: **V. Oza**[1], B. N. Lasseigne[2]; [1]Cell, Dev.al and Integrative Biology, UAB SOM, Birmingham, AL, [2]UAB SOM, Birmingham, AL
Disclosure Block: **V. Oza:** None.
Large scale cancer projects such as The Cancer Genome Atlas (TCGA) have generated multi-dimensional genomic data sets capturing orthogonal as well as complementary information across various cancers. To leverage the advantage of the available data, various integrative analytical approaches have been developed. However, such meta-dimensional analysis is challenging due to high dimensionality, high variability due to noise, and collinearity between different genomic features. Additionally, such analysis is complicated by the need for optimization of different analytical steps for integrating various genomic data. Although previous studies have focused on multi-omics analysis, the methods have not been optimized for individual steps in the multi-omics methods such as data standardization, data selection, feature reduction, machine learning classifier algorithms, and validation methods. Here, we develop an optimal framework for analysis of multi-omics datasets including gene expression, Copy Number Variation (CNV), and microRNA expression in ovarian cancer (as a proof of concept) to predict survival using TCGA data. We look at how the selection of data standardization, feature reduction, machine learning model, and cross-validation methods affect the multi-omics classifier performance in heterogeneous ovarian cancer data.

PrgmNr 3240 - Characterization and prediction of DNA methylation instability across human cancers

View session detail

Author Block: **S. Thalluri**, B. N. Lasseigne; Univ. of Alabama at Birmingham, SOM, Birmingham, AL
Disclosure Block: **S. Thalluri:** None.

DNA methylation instability (DNAmIN) occurs when there are globally altered DNA methylation (DNAm) signatures. While these patterns occur throughout the genome, extensive research to date in the cancer field has implicated that methylation changes occurring at CpG islands may be important biomarkers of disease etiology, progression, or treatment response. While many studies have examined this phenomenon, known as the CpG Island Methylator Phenotype, few have studied the relationship between non-CpG Island methylation, including CpG shores and CpG shelves, within and across cancers and with respect to clinical characteristics. We obtained DNAm array (450k Infinium Chip) data from over 11,000 patients in 33 cancers publicly available from The Cancer Genome Atlas (TCGA) database and compared and contrasted these DNAmIN metrics. In this study, we calculate different DNAmIN metrics and compare them to gain insight into the correlation between these signatures within and across cancer types. Our findings provide insights for generating hypotheses and contributing to the clinical relevance of DNAm as potential targets for future therapeutic intervention.

PrgmNr 3241 - Diagnosing Li-Fraumeni syndrome from the somatic genome

View session detail

Author Block: **B. Laverty**[1], V. Subasri[2], N. Light[1], D. Malkin[3]; [1]Univ. of Toronto, Toronto, ON, Canada, [2]Markham, ON, Canada, [3]Hosp for Sick Children, Toronto, ON, Canada
Disclosure Block: **B. Laverty:** None.
Li-Fraumeni syndrome (LFS) is a hereditary cancer predisposition syndrome caused by germline mutations of the tumour suppressor gene *TP53*. LFS is estimated to occur in 1:500 - 1:5000 people and is associated with an 80% lifetime cancer risk. LFS is diagnosed using personal and family cancer history and germline *TP53* sequencing. Several factors contribute to under-diagnosis, two of which I am going to tackle: germline *TP53* variants of uncertain significance (VUS), and the absence of germline TP53 mutations (30% of â□□classicâ□□ LFS patients).
We hypothesize that LFS cancers evolve uniquely from sporadic cancers, meaning the somatic genomes exhibit distinct characteristics that can diagnose the predisposition syndrome. To investigate this, we interrogated the mutational signatures, *TP53* copy number, *TP53* loss of heterozygosity, ploidy, and the incidence of chromothripsis in individuals with germline *TP53* mutations (n=27), somatic *TP53* mutations (n=17) and wild type for *TP53* (n=158). The proportion of single base substitution signatures 2, 5, 8, 13, 18 were significantly different between the groups as were ploidy, the incidence of chromothripsis and loss of heterozygosity. However, unsupervised hierarchical clustering of patients by these features did not completely separate LFS patients from non-LFS.
Next, we will analyze double base and indel mutational signatures and combine these features to train a logistic regression model with elastic net regularization using 10-fold cross-validation. We believe a machine learning algorithm may recognize patterns with all variants that are unique to LFS. This work will create a tool to diagnose two subsets of LFS patients that are currently hard to recognize. Early diagnosis of LFS facilitates surveillance for early detection of secondary tumours, leading to improved outcomes.

PrgmNr 3242 - Epigenomic landscape and 3D genome structure in pediatric high-grade glioma

View session detail

Author Block: **J. Wang**, T. Huang, Y. Hou, E. T. Bartom, X. Lu, A. Shilatifard, F. Yue, A. Saratsis; Northwestern Univ., Chicago, IL
Disclosure Block: **J. Wang:** None.
Pediatric high-grade gliomas (pHGGs), including glioblastoma multiforme (GBM) and diffuse intrinsic pontine glioma (DIPG), are morbid brain tumors. Even with treatment survival is poor, making pHGG the number one cause of cancer death in children. Up to 80% of DIPGs harbor a somatic missense mutation in genes encoding histone H3. To investigate whether H3K27M is associated with distinct chromatin structure that alters transcription regulation, we generated the first high-resolution Hi-C maps of pHGG cell lines and tumor tissue. By integrating transcriptome (RNA-seq), enhancer landscape (ChIP-seq), genome structure (Hi-C), and chromatin accessibility (ATAC-seq) datasets from H3K27M and wild-type specimens, we identified tumor-specific enhancers and regulatory networks for known oncogenes. We identified genomic structural variations that lead to potential enhancer hijacking and gene coamplification, including A2M, JAG2, and FLRT1. Together, our results imply three-dimensional genome alterations may play a critical role in the pHGG epigenetic landscape and contribute to tumorigenesis.

PrgmNr 3243 - MDC1 restrains ATR-mediated resection of DNA double-strand breaks in human cells

View session detail

Author Block: **S. S. Meyn**[1,2,3], P. S. Bradshaw[2]; [1]Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, [2]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada, [3]Dept. of Pediatrics, Univ. of Wisconsin - Madison, Madison, WI
Disclosure Block: **S.S. Meyn:** Receipt of Intellectual Property Rights/Patent Holder; Gene42.
Double strand DNA breaks (DSBs) are a potentially lethal form of cellular DNA damage that frequently cause genomic instability. Human cells primarily repair induced DSBs via the error prone mechanisms of c-NHEJ (classical-Non-Homologous End Joining), which occurs throughout the cell cycle, and Homologous Recombination (HR) repair, an error free process restricted to S and G2 phases that uses long stretches of homologous sequences. Defects in HR repair genes are responsible for the bone marrow failure syndrome Fanconi anemia, confer high risks for cancer, and can render tumors susceptible to PARP1-based therapies.
Pathway choice exerts a critical effect on the fidelity of DSB repair in human cells with resection determining whether the cell commits to c-NHEJ, or HR. In the G2 phase of the cell cycle, HR-mediated repair of induced DNA damage requires the protein product of the *ATM* gene, whose loss of function can cause the multisystem genetic disorder Ataxia-Telangiectasia.
We verify that, in human cells, ATM plays a major role in CtIP-mediated DSB resection and the creation of 3â□□ single-stranded DNA (ssDNA) ends used for HR. However, we find that DNA end resection in G2 can also be regulated independently of ATM by the cell cycle checkpoint protein MDC1. In human cells, MDC1 loss promotes unrestrained resection of DNA ends, as indicated by increases in the number and intensity of RPA-coated ssDNA foci in a process that is dependent upon ATR and CtIP. Further, optimal loading of the recombination protein Rad51 onto RPA coated ssDNA requires MDC1. While ATM promotes HR repair by initiating CtIP dependent DNA end resection, MDC1 can affect pathway choice through restraining HR repair by preventing ATR- and CtIP-dependent resection. Our work identifies a role for MDC1 in the control of DNA end resection that is independent of ATM and sheds mechanistic insight into the processing of DNA damage that is required for faithful DNA repair, maintenance of genome stability and prevention of tumorigenesis.

PrgmNr 3244 - Single cell transcriptome analysis maps aneuploidy to mesenchymal phenotypes during thyroid cancer progression

View session detail

Author Block: **L. Lu**[1], J. Wang[2], H. Ying[2], S. Bai[3], Y. Yan[4], C-k. Shiau[1], R. Kieser[1], M. Hu[3], X. Zhao[3], J. Wang[2], M. Williams[5], M. Zafereo[2], N. Navin[3], S. Lai[2], R. Gao[1]; [1]The Ctr. for Bioinformatics and Computational Biology, Houston Methodist Res. Inst.s, Houston, TX, [2]Dept. of Head and Neck Surgery, The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [3]Dept. of Genetics, The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [4]Vivian L. Smith Dept. of Neurosurgery, The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [5]Dept. of Pathology, The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX

Disclosure Block: **L. Lu:** None.

Anaplastic thyroid cancer (ATC) is arguably the most lethal human malignancy, with a median survival of ~12 weeks.About half ATC tumors have prior differentiated forms of thyroid cancer types, commonly papillary thyroid cancer (PTC), suggesting an evolutionary continuum from PTC to ATC. In comparison to the extreme lethality of ATC, PTC patients are typically indolent with long life-histories. The molecular mechanisms driving ATC progression have been challenging to resolve, mostly due to the heterogenous cellular systems that are undergoing dynamic changes and the rarity of human ATC tissue samples. In this study, we applied high throughput 3â□□ scRNAseq to unbiasedly analyze the cell populations in thyroid tumors. In total, we analyzed 79,747 single cell transcriptomes from 11 ATC tumors, 7 PTC tumors and 7 normal thyroids. To make the precise inference of tumor cell lineages, we stratified the epithelial cells into 5 categories by calculating genotypes from 3â□□ sc-RNAseq data that include normal epithelial cells, two subtypes of PTC cancer cells and two subtypes pf ATC cancer cells (i.e. inflammatory ATC and mesenchymal ATC). Our data showed that aneuploidy is strongly associated with mesenchymal phenotypes in ATC cancer cells, whereas inflammatory ATC cancer cells often only have 2N copy numbers. Mesenchymal ATCs are characterized with co-expression of epithelial and mesenchymal phenotypes and loss of thyroid differentiation programs. Several abnormal mitotic checkpoint pathways associated with survival of aneuploids are enriched in mesenchymal ATC cells. We further analyzed the transcriptional programs of immune] and stromal cell types. Our results showed that T cells in ATC are more exhausted while macrophages undergo M2 polarization and mesenchymal transition. In addition, we found that ATC tumors are enriched towards inflammatory cancer associated fibroblasts (CAFs) that also have strong mesenchymal features. A significantly tight relationship between these mesenchymal cell types (i.e. cancer cells, myeloid cells and CAFs) are defined through intercellular network reconstruction. Collectively, our results suggest that ATC progression is driven by aneuploidy associated mesenchymal transitions in both cancer and non-cancer cell types, providing novel insights into targeting this extremely lethal disease.

PrgmNr 3245 - Statistical analysis of breast cancer co-methylation patterns

View session detail

Author Block: **S. Sun**[1], J. Dammann[2], P. Lai[3], C. Tian[4]; [1]Texas State Univ., San Marcos, TX, [2]St. Stephenâ□□s Episcopal Sch., Austin, TX, [3]Massachusetts Inst. of Technology, Cambridge, MA, [4]Liberal Arts and Sci. Academy, Austin, TX
Disclosure Block: **S. Sun:** None.
Breast cancer is one of the most commonly diagnosed cancers. It is associated with DNA methylation, which is an epigenetic event. In an adult human genome, methylation occurs when a methyl group is covalently added to a cytosine base that is paired with a guanine, i.e., a CG site. DNA methylation can have different effects on gene functions. The methylation patterns of different genes or sites in a genome may be correlated in certain ways. This correlation pattern is known as co-methylation. It is still not clear how different genes co-methylate with each other in the whole genome of breast cancer samples. Two previous studies are conducted based on relatively small datasets, Illumina methylation 27K data with only 27,000 CG sites per sample. In this study, we analyze much larger publicly available datasets, Illumina methylation 450K (i.e., 450,000 sites) data of 53 breast cancer patients including their matched normal data. Our key findings are summarized below. First, normal samples have 1.7 times more highly correlated or co-methylated CG pairs than tumor samples. Both tumor and normal have more than 93% of positive co-methylation, but normal samples have significantly more negatively correlated CG sites than tumor samples (6.6% vs 2.8%). Second, both tumor and normal samples have about 94% of co-methylated CG pairs on different chromosomes, but normal samples have 470 million more CG pairs. For the highly co-methylated pairs on the same chromosome, they tend to be located close to each other. Third, a small proportion of CG sitesâ□□ co-methylation patterns change dramatically from normal to tumor. The percentage of differential methylation (DM) sites among them is larger than the DM rate in the whole genome (10.53% vs 8.56%). Fourth, certain CG sites are highly co-methylated with other CG sites; the top 100 of such CG sites in tumor and normal samples have no overlaps. Fifth, chromosome X co-methylation patterns are very different from other chromosomes. Finally, the network analyses of genes associated with several sets of co-methylated CG sites mentioned above show that tumor and normal samples have different patterns as well. Our findings will provide researchers with a new and improved understanding of co-methylation patterns in breast cancer. Our ability to thoroughly analyze co-methylation of large datasets will allow researchers to study relationships and associations between different genes in breast cancer.

PrgmNr 3246 - Subtype-associated epigenomic landscape and 3D genome structure in bladder cancer

View session detail

Author Block: **Q. Wang**[1], T. Iyyanki[1], B. Zhang[1], Y. Hou[1], Q. Jin[1], J. Xu[1], H. Yang[1], T. Liu[1], X. Wang[1], F. Song[1], Y. Luan[1], H. Yamashita[2], R. Chien[3], H. Lvu[1], L. Zhang[2], L. Wang[1], J. Warrick[2], J. Raman[2], J. Meeks[1], D. DeGraff[2], F. Yue[1]; [1]Northwestern Univ., Chicago, IL, [2]The Pennsylvania State Univ., Hershey, PA, [3]Univ. of Illinois, Chicago, IL

Disclosure Block: **Q. Wang:** None.

Muscle-invasive bladder cancers are characterized by their distinct expression of luminal and basal genes, which could be used to predict key clinical features such as disease progression and overall survival. Transcriptionally, FOXA1, GATA3, and PPARG are shown to be essential for luminal subtype-specific gene regulation and subtype switching, while TP63, STAT3, and TFAP2 family members are critical for regulation of basal subtype-specific genes. Despite these advances, the underlying epigenetic mechanisms and 3D chromatin architecture responsible for subtype-specific regulation in bladder cancer remain unknown. We determine the genome-wide transcriptome, enhancer landscape, and transcription factor binding profiles of FOXA1 and GATA3 in luminal and basal subtypes of bladder cancer. Furthermore, we report the first-ever mapping of genome-wide chromatin interactions by Hi-C in both bladder cancer cell lines and primary patient tumors. We show that subtype-specific transcription is accompanied by specific open chromatin and epigenomic marks, at least partially driven by distinct transcription factor binding at distal enhancers of luminal and basal bladder cancers. Finally, we identify a novel clinically relevant transcription factor, Neuronal PAS Domain Protein 2 (NPAS2), in luminal bladder cancers that regulates other subtype-specific genes and influences cancer cell proliferation and migration.In summary, our work identifies unique epigenomic signatures and 3D genome structures in luminal and basal urinary bladder cancers and suggests a novel link between the circadian transcription factor NPAS2 and a clinical bladder cancer subtype.

PrgmNr 3247 - Survival analysis and evaluation of germline genomic patterns across multiple cancers using boosted trees and model interpretability

View session detail

Author Block: **L. Chapman**[1,2], Z. Boukouvalas[2]; [1]NIH, Rockville, MD, [2]American Univ., Washington, DC

Disclosure Block: **L. Chapman:** None.

Approximately 5-10% of all cancers can be linked to heritable genetic variants. Evaluation of germline genetic variants in cancer has largely been limited to variants within a certain set of genes known to be associated with cancer. Previous studies have shown that approximately 8% of the 10389 patients from The Cancer Genome Atlas (TCGA) carried pathogenic germline variants. Common variants have smaller effect sizes; however, many select variants could collectively contribute to cancer development. In the proposed study, we evaluate germline variants that are associated with survival across multiple cancer types in TCGA. We apply a boosted tree method - XGBoost - to select a feature set consisting of non-linear SNP interactions within each cancer type, and classify patients based on survival using SVM. Over 1 million germline substitutions from TCGA representing multiple cancer types will be evaluated. To identify sets of germline SNPs that are significantly associated with cancer survival predictions, model interpretation is conducted using a local surrogate model: local interpretable model-agnostic explanations (LIME). Spearman correlation is used to identify associations between SNP burdens and COSMIC somatic signatures to evaluate the possible impact of germline SNP networks on oncogenic pathways. Our approach will enable us to identify putative therapeutic targets as well as possible markers for survival will be identified, and could also possibly discover clinically actionable non-coding genetic variants within the networks.

PrgmNr 3248 - The clinical impact of somatic variants in *AGT*, *MGMT* and *TP53* in Mexican patients with astrocytoma

View session detail

Author Block: **J. Carlos-Escalante**[1], T. Wegman-Ostrosky[2]; [1]Inst. Natl. de CancerologÃa, Jesus Maria, Mexico, [2]NCI, Rockville, MD
Disclosure Block: **J. Carlos-Escalante:** None.
Introduction: Astrocytomas are the most common primary malignant brain tumor in adults. They are highly disabling, and the prognosis of patients is dismal, particularly for grade IV astrocytomas. Clinical prognostic factors for astrocytomas include sex, age, performance status, extent of resection, and adjuvant chemoradiotherapy. Molecular prognostic factors have been described, being the presence of somatic mutations in *IDH1* or *IDH2* the best characterized. Somatic mutations in other genes could serve as molecular prognostic factors as well. *AGT* encodes for angiotensin, which is a central piece of the renin-angiotensin system (RAS). Somatic mutations in *AGT* could be related to astrocytoma prognosis as most components of the RAS are expressed in human astrocytomas. *MGMT* promoter methylation is an already established predictive marker for astrocytoma therapy, however, the role of somatic mutations in this gene in prognosis is unknown. *TP53* is commonly mutated in astrocytomas, nonetheless, the prognostic value of these mutations is still debated.
Objective: The aim of this work was to evaluate the association of variants in *AGT*, *MGMT*, and *TP53* with astrocytoma prognosis.
Materials and Methods: Tumor tissue and peripheral blood samples from 48 patients with astrocytoma from the Instituto Nacional de NeurologÃa y NeurocirugÃa (INNN) were analyzed through targeted panel sequencing for *AGT*, *MGMT*, and *TP53.* Somatic mutations called were filtered and annotated. The data from 503 individuals with astrocytoma from TCGA were analyzed as validation cohort.
Results: 61 mutations were detected in the samples from Mexican patients. *TP53* mutations associated to loss of function were associated with higher 2-year survival in our cohort (OR 4.14, 95% CI: 1.02-16.8, p=0.039). *TP53* mutations in general were associated with longer overall survival after adjustment in the TCGA cohort (HR: 0.71, 95% CI: 0.5-0.99, p=0.049). Women in our cohort tended to have more mutations than men, without reaching statistical significance, however, mutations in *AGT* and *MGMT* were found only in females. *AGT* mutations were present only in females of TCGA cohort.
Conclusions: In this study we found that the presence of *TP53* somatic mutations may be a prognostic factor for patients with astrocytoma. In addition, we observed a sex-differentiated mutation distribution for *AGT*.

PrgmNr 3249 - Tumor spectrum of *FLCN* gene mutation carriers in an unselected pan-cancer patient population

View session detail

Author Block: **S. Kane**[1], Y. Kemel[2], A. Maio[1], N. Mehta[3], D. Mandelker[3], Z. Stadler[1], M. I. Carlo[1]; [1]Clinical Genetics, Med., Mem. Sloan Kettering Cancer Ctr., New York, NY, [2]Sloan Kettering Inst., Mem. Sloan Kettering Cancer Ctr., New York, NY, [3]Pathology, Mem. Sloan Kettering Cancer Ctr., New York, NY

Disclosure Block: **S. Kane:** None.

Background: Birt-Hogg-DubÃ© syndrome (BHD) is a genetic syndrome caused by deleterious germline mutations in the *FLCN* gene. BHD is characterized by benign skin lesions, pulmonary cysts and pneumothorax, and increased risk of renal tumors. Whether the phenotype extends beyond these features is unknown, as most patients with BHD are identified when they present with classic features. We describe the phenotypes of patients with BHD from an unselected, pan-cancer population.

Methods: 22,440 patients with cancer were tested via MSK-IMPACT, a matched tumor-normal next generation sequencing platform. Germline testing included >70 genes associated with cancer risk. Patients with deleterious *FLCN* germline mutations were identified. Clinical features were obtained from medical records and clinical notes. Tumor sequencing was analyzed for loss of heterozygosity (LOH) using the Fraction and Allele-Specific Copy Number Estimates from Tumor Sequencing (FACETS) algorithm.

Results: 12 patients (0.05%) had deleterious germline mutations *FLCN*. Median age at cancer diagnosis was 59 (range 22-73). Cancers included RCC (n=2), colorectal (n=2), prostate (n=2), and 1 each with breast, thyroid, pancreas, uterus, undifferentiated pleomorphic sarcoma, squamous cell carcinoma (skin) and non-small cell lung cancer. One patient had 3 cancers: synchronous colon and rectal, and prostate cancers. Of the 2 patients with RCC, pathology was â□□unclassifiedâ□□ RCC. Of 6 patients with available colonoscopy records, 5 (83%) had colon polyps. 9 patients (75%) had emphysematous changes on imaging. 6 patients (50%) had no personal or family history of BHD features noted prior to genetic testing. At least 3/12 tumors (breast, kidney, and sarcoma; 25%) were found to have *FLCN* LOH.

Conclusions: BHD is a rare genetic syndrome and the spectrum of associated tumors is broad. The phenotypic spectrum of patients with *FLCN* gene mutations from this pan-cancer population differ from patients described in the literature, many of which were tested due to features consistent with BHD. Most patients in this study were found on imaging to have emphysematous lung changes, which may correspond to the cystic lung findings seen in BHD. Identification of *FLCN* mutations can direct cancer screening, lifestyle modification and cascade testing in at-risk relatives. Of interest, several non-canonical cancers were found to have LOH, suggesting a possible role for *FLCN* in the pathogenesis of the tumor. Our study suggests that criteria for BHD evaluation could be broadened due to the clinical variability of patients identified in this study. The role of *FLCN* in the development of other cancers can be further explored.

PrgmNr 3250 - Up-regulation of chromatin remodeling genes contributes to chemoresistance and tumorigenesis in PDAC

View session detail

Author Block: **C. Wright**[1,2], E. R. Gordon[1], R. M. Myers[1], S. J. Cooper[1]; [1]HudsonAlpha Inst. for Biotechnology, Huntsville, AL, [2]The Univ. of Alabama in Huntsville, Huntsville, AL
Disclosure Block: **C. Wright:** None.
Pancreatic ductal adenocarcinoma (PDAC) is the most common form of pancreatic cancer and one of the most lethal forms of cancer with a five-year survival rate of 10%. Existing therapies are not sufficient and the few recurrent mutations found in PDAC tumors have failed to yield effective targeted therapies. The combination of late-stage diagnosis, lack of targeted therapies and resistance to existing therapies contribute to poor outcomes. We used genome wide CRISPR screens to show that activation of chromatin remodeling genes including *HDAC1* appears to contribute to multi-drug resistance. HDAC1 overexpression is also correlated with poor patient outcome. Transcriptomic analysis shows that HDAC1 overexpression may lead to drug resistance by inducing cell state transition. We generated PDAC cell lines that we endogenously activated HDAC1 expression using the dCas9-VP64 epigenome editing system. ChIP-sequencing and RNA-sequencing were performed on HDAC1 activated cell lines, control cell lines, and a normal pancreas cell line to find novel HDAC1 binding sites and altered chromatin accessibility, measured by H3K27ac ChIP-seq, that are associated with gene expression changes previously linked to chemoresistance. We compared the presence of H3K27ac between the cell lines to identify active promoters and enhancers that are regulated by HDAC1 and contribute to drug resistance. Combining DNA binding site data with gene expression data following HDAC1 overexpression showed how new binding sites alter gene expression programs that result in a more stem-like state and contribute to chemoresistance in pancreatic cancer. An improved understanding of the regulatory elements that are relevant for resistance might help us understand how mutations contribute to drug resistance in pancreatic cancer. Our work provides an improved understanding of the regulatory landscape in PDAC, and will particularly resolve the genomic changes most relevant to drug resistance with the goal of inspiring new treatment options.

PrgmNr 3251 - Assessment of population-appropriate polygenic risk scores for lipid traits in African Americans

View session detail

Author Block: **D. Drouet**, S. Liu, D. C. Crawford; Case Western Reserve Univ., Cleveland, OH
Disclosure Block: **D. Drouet:** None.
Polygenic risk score (PRS) studies for cardiovascular disease (CVD), while promising, are based on genome-wide association studies in European-descent populations. Here we calculate population-appropriate PRS in an African American (AA) patient population from a biobank linked to electronic health records (EHRs) for risk factors of CVD: high density lipoprotein cholesterol (HDL-C), low density lipoprotein cholesterol (LDL-C), triglycerides (TG) and total cholesterol (TC). We limited analyses to individuals (n=1,898; 1,898; 1,978; and 2,106) with Illumina Metabochip genotypes available for 44, 36, 51, and 48 previously associated HDL-C, LDL-C, TG, and TC single nucleotide variants, respectively, reported by the Population Architecture using Genomics and Epidemiology (PAGE) Study. Using PAGE Study risk allele and effect estimates, we calculated an unweighted and weighted PRS for each individual and tested each for an association with each lipid trait using linear regressions unadjusted and adjusted for age, sex, and body mass index. The majority of patients (63.4%) are female and are on average (46.20) middle age. Almost one-third had mention of heart attack keywords in their EHR problems list; 1% and 1.5% had current procedural terminology (CPT) codes for coronary artery bypass graft and angioplasty and/or coronary stents, respectively, and 1.7% had three mentions of International Classification of Diseases (ICD)-9 codes for myocardial infarction. Mean weighted PRS was 3.44, 2.2, 5.96, and 3.69 for HDL-C-, LDL-C, TG, and TC, respectively, in this AA patient population. Despite previous reports of strong associations between each lipid trait and genetic load, weighted PRS (adjusted betas = -0.65, -1.52, -6.72, and -0.86, respectively) was not associated with HDL-C, LDL-C, TG, or TC levels in this AA patient population (p-values>0.05). None of the PRS was associated with CVD; however, AA patients in the 90th percentile of genetic load trended towards higher number of CVD cases compared with those with less genetic load. These data suggest this population appropriate PRS for these lipid traits cannot predict risk for CVD in AA, likely due to the small sample sizes, small individual genetic effect sizes, and/or the need for additional variants to be included in the PRS.

PrgmNr 3252 - Bivariate genome-wide association study of circulating fibrinogen and C-reactive protein (CRP) levels

View session detail

Author Block: **J. Hahn**[1], G. Temprano-Sagrera[2], N. L. Smith[3,4,5], S. Ligthart[6], A. Dehghan[7,8], M. Sabater-Lleal[2,9], A. C. Morrison[1], P. S. de Vries[1]; [1]Human Genetics Ctr., Dept. of Epidemiology, Human Genetics, and Environmental Sci., Sch. of Publ. Hlth., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Genomics of Complex Diseases, Res. Inst. of Hosp. de la Santa Creu i Sant Pau, IIB Sant Pau, Barcelona, Spain, [3]Dept. of Epidemiology, Univ. of Washington, Seattle, WA, [4]Kaiser Permanente Washington Hlth.Res. Inst., Seattle, WA, [5]Seattle Epidemiologic Res. and Information Ctr., Dept. of VA Office of Res. and Dev., Seattle, WA, [6]Dept. of Epidemiology, Erasmus Univ. Med. Ctr., Rotterdam, Netherlands, [7]UK Dementia Res. Inst. at Imperial Coll. London, London, United Kingdom, [8]Dept. of Epidemiology and Biostatistics, Imperial Coll. London, London, United Kingdom, [9]Cardiovascular Med. Unit, Dept. of Med. Solna, Karolinska Inst.t, Ctr. for Molecular Med. and Karolinska Univ. Hosp. Solna, Stockholm, Sweden

Disclosure Block: **J. Hahn:** None.

Introduction: Previous genome-wide association studies (GWAS) have identified 41 loci associated with circulating fibrinogen and 58 loci with C-reactive protein (CRP) levels. Both proteins play an important role in inflammatory pathways, and several of the previously described loci overlap between fibrinogen and CRP, suggesting that they share a common genetic background. Hence, by integrating these phenotypes in a bivariate GWAS, we aimed to identify novel genetic variants that are pleiotropic and associated with both fibrinogen and CRP. **Methods**: We performed a bivariate GWAS by combining summary statistics of previously published GWAS on fibrinogen (n=120,246) from the CHARGE Consortium and CRP (n= 363,228) from UK Biobank, both comprised of European individuals, using metaUSAT, a unified score-based association test. Loci were considered to be pleiotropic when: 1) the bivariate p-values were 5Results: We identified 81 pleiotropic loci, of which 51 were novel for fibrinogen, and 14 were novel for CRP. Of these, 11 loci were novel for both fibrinogen and CRP. Out of the 14 novel pleiotropic loci for CRP, using Bonferroni correction of 3.57E-3 (0.05/14), 7 genetic variants were replicated. They were located in or near genes *CCT3*, *MICU1*, *ARL14EP*, *GSDMA*, *NPEPL1*, *NR1I2*, and *RNA5SP158*. **Conclusion**: Bivariate GWAS was an efficient method to identify numerous pleiotropic loci for fibrinogen and CRP. Our study identified 81 pleiotropic loci associated with both fibrinogen and CRP. From these, we discovered and replicated 7 new loci for CRP. This shows that fibrinogen and CRP indeed share a common genetic architecture with many pleiotropic loci.

PrgmNr 3253 - Blood Pressure Polygenic Score Associations in a Pediatric Cohort Requiring Surgery for Congenital Heart Defects

View session detail

Author Block: **J. Breeyear**[1], J. M. Keaton[2], A. H. Smith[3], D. Klarin[4], S. M. Damrauer[5], P. Natarajan[4], S. L. Van Driest[6], J. G. Weiner[6], P. J. Kannankeril[6], T. L. Edwards[6]; [1]Vanderbilt Univ., Nashville, TN, [2]NIH/NHGRI, Clarksburg, MD, [3]Johns Hopkins All Children's Hosp., St. Petersburg, FL, [4]Massachusetts Gen. Hosp., Boston, MA, [5]Hosp. of the Univ. of Pennsylvania, Philadelphia, PA, [6]Vanderbilt Univ. Med. Ctr., Nashville, TN

Disclosure Block: **J. Breeyear:** None.

Congenital malformations are the leading cause of infant mortality in the United States. Congenital heart defects (CHD) are the most prominent congenital malformation, affecting 40,000 US births per year: with most children requiring surgical intervention(s). Better understanding of the basis for clinical variability in morbidity and mortality has the potential to improve clinical management and outcomes. To date, few studies have evaluated blood pressure polygenic risk scores (PRS) associations with post-operative outcomes. We used imputed genotypes from pediatric participants requiring surgery for CHD (median age = 201 days, nmax = 2,498). Base data for systolic (SBP) and diastolic (DBP) blood pressure PRSs used published GWAS data (nmax = 760,226). The PRSs were validated in BioVU (nmax = 37,132) with genetic data pruned for linkage disequilibrium at an r2 threshold of 0.1 at a maximum distance of 250 kilobases from associated SNPs in the BP summary statistics. PRSs associations with in-hospital mortality (HM), intensive care unit length of stay (ICU-LOS), length of hospital stay (LOS), and vasoactive-inotropic scores (VIS) were modeled using logistic or linear regression in STATA, adjusted for age, sex, BMI, Society for Thoracic Surgeons-European Association for Cardio-Thoracic Surgery category, and the top ten principal components of ancestry. The DBP PRS was associated with decreased HM risk (0.64 (0.46 - 0.89), p = 0.0085), decreased ICU-LOS (-1.79 Â± 0.66 days, p = 0.0071), decreased LOS (-2.16 Â± 0.91, p = 0.018), and decreased VIS (-0.52 Â± 0.23, p = 0.022) The SBP PRS was not significantly associated with any outcomes. A combined model of the DBP and SBP PRSs was associated with decreased HM risk (0.48 (0.29 - 0.78), p = 0.0033), decreased ICU-LOS (-2.36 Â± 0.61 days, p = 0.000096), and decreased LOS (-3.74 Â± 1.46, p = 0.011). Our findings present evidence that pediatric patients with a genetic predisposition to high blood pressure have a reduced risk of in-hospital mortality and a protective effect in pediatric post-surgical recovery following congenital heart surgery. The observed associations of blood pressure PRSs with post-operative outcomes provides insight into the use of adult-derived cardiovascular PRSs as tools to anticipate negative outcomes and understand inter-individual variability of response to CHD repair in children.

PrgmNr 3254 - Detecting blood-pressure regulatory mechanisms with tissue specific CRE maps and biobank scale datasets

View session detail

Author Block: **O. Yaacov**[1], D. Lee[2,1], P. Mathiyalagan[1], X. Zhu[3], S. Ganesh[4], A. C. Morrison[5], N. J. Risch[6], A. Chakravarti[1]; [1]Ctr. for Human Genetics and Genomics, NYU Grossman Sch. of Med., New York, NY, [2]Boston Children's Hosp., Boston, MA, [3]Case Western Reserve Univ, Cleveland, OH, [4]Univ. of Michigan, Ann Arbor, MI, [5]Univ. of Texas at Houston, Houston, TX, [6]Univ. of California San Francisco, San Francisco, CA

Disclosure Block: **O. Yaacov:** None.

Blood pressure (BP) is a common quantitative trait with ~30% heritability. Numerous genetic studies have clearly demonstrated that the majority of this heritability is explained by ~900 genetic associations of small effect (median effect size ~0.25â□□mm Hg per allele) distributed across the genome. In this study, we demonstrate how genome-wide epigenomic maps can be used to identify the contributions of specific genes and variants in the kidney, heart, artery and adrenal gland, to BP regulation. We collected tissue samples from human kidneys, hearts, arteries and adrenal glands and performed both ATAC- and RNA- sequencing. Combined with H3K27ac and ATAC-seq data from the ENCODE project we created tissue-restricted open chromatin maps, using our prior published methods. We used these maps to train machine learning models, delta-SVM, an established method to predict tissue-specific CRE variants. We scored all common (minor allele frequency>1%) single nucleotide variants (SNVs) from the 1000G project in each of the tissue -restricted accessible chromatin peaks. We then performed a custom SKAT gene-wise analysis, using the delta-SVM scores as weights and limiting analysis to SNVs within chromatin peaks. For phenotypes, we used systolic (SBP) and diastolic (DBP) BP measurements (corrected for age, BMI, sex and anti-hypertensive medication use) and imputed genotypes of 77,823 individuals from the RPGEH biobank (Research Program on Genes, Environment, and Health, Kaiser Permanente North California). We Identified 309, 259, 331 and 367 statistically significant genes (Î±=0.05 BH corrected) for DBP in the tibial artery, kidney, heart left ventricle (LV) and adrenal gland, respectively. Similarly, we detected 191, 184, 204 and 204 genes for SBP. 50-70% of these genes were replicated using an independent dataset of 315,270 unrelated individuals from the UK Biobank. This is highly significant because permutation analysis of these data replicated

PrgmNr 3255 - Epigenetic age acceleration predicts subclinical atherosclerosis among participants of the Bogalusa Heart Study

View session detail

Author Block: **X. Sun**, W. Chen, A. C. Razavi, M. Shi, E. Harville, J. He, L. A. Bazzano, T. N. Kelly; Tulane Univ., New Orleans, LA
Disclosure Block: **X. Sun:** None.
Background: Epigenetic age acceleration (EAA), or increased DNA methylation (DNAm)-based age relative to chronological age, has been associated with cardiovascular disease (CVD) risk factors. However, the relationship between EAA and subclinical atherosclerosis is unclear. We examined associations of EAA with carotid intima-media thickness (cIMT) and carotid plaque in the Bogalusa Heart Study (BHS).
Methods: We included 1,487 BHS participants with simultaneous measures of whole-blood DNAm and cIMT. DNAm data generated using the Illumina HumanMethylation450 BeadChip was used to estimate extrinsic EAA (EEAA) and intrinsic EAA (IEAA), based on Hannumâ□□s and Horvathâ□□s epigenetic clocks, respectively. cIMT was measured using carotid ultrasound, with carotid plaque defined as a distinct focal wall thickening of >1.5 mm. Cross-sectional associations of EAA with cIMT and carotid plaque were tested using multiple linear and logistic regression models, respectively. Three multivariable models were employed, adjusting for: age, sex, and race (model 1); model 1 covariables plus education, smoking, and drinking (model 2); and model 2 covariables plus blood pressure, lipids, and glucose (model 3). Among 321 BHS participants with an additional prospective measure of cIMT, we used logistic regression to test whether baseline EAA predicted incident carotid plaque over an average of 6.0 years follow-up. The same multivariable models were employed, additionally adjusting for baseline cIMT and follow-up time in all models.
Results: Individuals were on average 47 years old, 56% were women, and 33% were African American. Each standard deviation increase in EEAA was cross-sectionally associated with a 0.037 mm increase in cIMT in model 1 ($P=3.09Ã□10^{-6}$), a 0.031 mm increase in model 2 ($P=8.35Ã□10^{-5}$), and a 0.021 mm increase in model 3 ($P=5.97Ã□10^{-3}$). Similarly, each standard deviation increase in EEAA was cross-sectionally associated with a 1.29-fold higher odds of carotid plaque in model 1 ($P=6.66Ã□10^{-5}$), a 1.23-fold higher odds in model 2 ($P=1.24Ã□10^{-3}$), and a 1.17-fold higher odds in model 3 ($P=2.3Ã□10^{-2}$). In prospective analyses, EEAA significantly predicted development of carotid plaque in model 1, with similar but marginally significant trends observed in models 2 and 3. EEAA was associated with a 1.32-fold higher odds of incident carotid plaque in model 1 ($P=3.01Ã□10^{-2}$), 1.28-fold higher odds in model 2 ($P=6.12Ã□10^{-2}$), and 1.24-fold higher odds in model 3 ($P=7.76Ã□10^{-2}$). IEAA was not associated with cIMT or carotid plaque.
Conclusions: We identified significant associations of EEAA with subclinical atherosclerosis, which were independent of known CVD risk factors.

PrgmNr 3257 - Identification of genetic variants that confer susceptibility to familial combined hyperlipidemia

View session detail

Author Block: **H. Ata Ul Mustafa**[1], R. K. Patel[1,2], S. Wasi Haider[1,3], K. M. Kunji[4], J. Fan[1], R. Turner[1], D. Lucas[5], H. Omar[6,7], W. Neal[8], E. A. Rosenthal[9], M. Saad[4], D. A. Primerano[1], A. Q. Nato[1]; [1]Dept. of BioMed. Sci., Joan C. Edwards Sch. of Med., Marshall Univ., Huntington, WV, [2]Tata Consultancy Services Limited, Edison, NJ, [3]Div. of Electrophysiology and Heart Failure, Dept. of Software Verification and Validation, Abbott Lab., Saint Paul, MN, [4]Qatar Ctr. for Artificial Intelligence, Qatar Computing Res. Inst., Hamad Bin Khalifa Univ., Doha, Qatar, [5]Charleston Area Med. Ctr. Hlth.Ed. and Res. Inst., Inc, Charleston, WV, [6]Adolescent Hlth.Ctr., Lehigh Valley Reilly Childrenâ□□s Hosp., Allentown, PA, [7]Dept. of Pediatric and Adolescent Med., Homer Stryker MD Sch. of Med., Western Michigan Univ., Kalamazoo, MI, [8]Dept. of Pediatrics, Sch. of Med., West Virginia Univ., Morgantown, WV, [9]Univ. of Washington Sch. of Med., Seattle, WA

Disclosure Block: **H. Ata Ul Mustafa:** None.

Familial combined hyperlipidemia (FCHL) is a common genetic disorder that is characterized by elevated cholesterol and triglycerides. The disorder exhibits diverse clinical manifestations with complex etiology. Our goal is to detect single nucleotide variants (SNVs) associated with FCHL within regions with evidence of linkage to the disease. We hypothesize that variants associated with FCHL may affect gene function or lipid metabolism and contribute to its etiology. FCHL families were ascertained by: (1) identifying probands with fasting plasma total cholesterol â□¥95th percentile and triglycerides â□¥90th percentile for age, sex, and race and (2) having at least 2 participants with FCHL and 1 unaffected relative. We had a total of 638 individuals (315 males, 304 females, 19 unknown) in 41 extended families (size: 4-46 individuals) from 3 centers in WV and KY. We excluded families who only had singletons, without DNA or only have 2-3 individuals with DNA. For families included in the analysis, we trimmed individuals to focus on pedigree branches having affected individuals who are closely related. After trimming, we have a total of 446 individuals (211 males, 218 females, 17 unknown) in 23 families (size=4-41) with ages ranging from 2 to 95 years old (Î¼= 40.51Â± 21.63years old). As for phenotype data, fasting plasma total cholesterol ranged from 94-490 mg/dL (Î¼= 226.14Â±64.47 mg/dL) and triglycerides ranged from 37-3021 mg/dL (Î¼=220.11Â±273.25 mg/dL). A total of 180 individuals were genotyped using ~654K SNP panel. Of these, we conducted whole exome sequencing (WES) in 119 individuals that were selected using GIGI-Pick to capture segregation of founder alleles. We use the pedigree-based analysis pipeline (PBAP) to: (1) perform file manipulations, (2) validate pedigree structures, (3) sub-select linkage disequilibrium (LD)-reduced panel of markers from the dense SNP array, and (4) access gl_auto of the MORGAN package to sample inheritance vectors (IVs), which specify the flow of founder alleles in the pedigree. Using sampled IVs, we perform linkage analyses using PBAP to identify genomic region(s) of interest (ROI) that have evidence of linkage to FCHL. Within ROI(s), we perform family-based genotype imputation in individuals without WES data that enables us to obtain genotypes at a much lower cost. We perform family-based association analysis to determine variants associated with FCHL using both observed and imputed genotypes.

PrgmNr 3258 - Phenotypic features of TGFB3 disease-causing variants in a group of North Americans

View session detail

Author Block: **N. Garg**[1], J. R. Schweber[2], J. Fleischer[1]; [1]Southern Illinois Univ. Sch. of Med., Springfield, IL, [2]Washington Univ. Sch. of Med., St Louis, MO
Disclosure Block: **N. Garg:** None.
TGFB3 gene mutations, newly described as a cause of Loeys-Dietz Syndrome type IV (LDS4) in 2015, have been reported in around 80 patients to date, with the largest cohorts being described by Marsili et al, 2020 and Schepers et al, 2018. To further delineate the phenotype associated with TGFB3 mutations, we present 7 new cases of LDS4 with two novel TGFB3 mutations with comparison to the previously reported cohorts. We also describe a patient with an FBN1 variant in combination with a TGFB3 variant and the possible effect of this variant combination on the phenotype.
In our cohort, we found joint hyperlaxity and pectus deformity were the most common features and were present in 71% of the cases. Scoliosis was the second most common finding present in 57% of the cases. Twenty-Eight percent were found to have arterial tortuosity. Aortic root aneurysm, ascending aortic aneurysm and mitral valve prolapse represented the most common cardiovascular findings, and each was present in 14% of the cases. Other prominent features included myopia in 28%, bifid uvula in 42%, flat feet in 42%, and stria in 28% of these individuals. In familial cases, incomplete penetrance and variable clinical expressivity were noted. In summary, TGFB3 variants had multisystemic involvement and were associated with a significant aortic disease as well as arterial tortuosity, highlighting the importance of early diagnosis of TGFB3 mutations to provide focused screening to prevent morbidity and mortality in such individuals.

PrgmNr 3259 - Polygenic risk for atrial fibrillation across ancestry: the Electronic Medical Records and Genomics (eMERGE) Network

View session detail

Author Block: **M. J. Puckelwartz**[1], K. T. Sawicki[1], L. L. Pesce[1], S. D. Kearns[1], E. R. Duvall[1], J. A. Pacheco[1], C. Liu[2], M. S. V. Elkind[2], N. Vaitinadin[3], G. E. Davogustto[3], M. B. Shoemaker[3], J. D. Mosley[3], L. J. Rasmussen-Torvik[1], R. L. Chisholm[1], D. M. Roden[3], E. M. McNally[1]; [1]Northwestern Univ., Chicago, IL, [2]Columbia Univ., New York, NY, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN
Disclosure Block: **M.J. Puckelwartz:** None.
Rationale: Atrial fibrillation (AF) is the most common cardiac arrhythmia, affecting over 30 million people worldwide. In the US, lifetime risk for AF differs across groups with Whites having the highest risk, followed by Blacks, and with Hispanic and Asians having the lowest incidence of AF. These lower risks occur in spite of higher occurrence of comorbidities that increase risk of AF in some non-White groups. The role of inherited susceptibility to AF is well established, from rare monogenic disorders to polygenic traits. Polygenic risk scores (PRS) utilize risk allele information to generate a single measure of inherited susceptibility. Genome-wide association studies (GWAS) have identified independent risk variants for AF, but these are derived largely from European ancestry cohorts and may not be applicable to non-European ancestries.
Objective: Determine if polygenic risk scores (PRS) derived from the AF risk variants have similar discrimination in non-European ancestries in the diverse Electronic Medical Records and Genomics (eMERGE) Network cohort.
Methods: The AF PRS was applied to a diverse cohort of 39,908 participants (self-reported: 89.3% European, 6.9% African and 2.7% Hispanic and 1.1% Asian) in eMERGE I-III with and without AF and genotype data. Diagnosis of AF was based on electronic health record (EHR) information including inpatient International Classification of Disease (ICD-9, ICD-10) diagnosis codes. The PRS was generated using 163 SNPs derived from two published GWAS studies with ~1.65 million subjects (97% European, 0.55% African, 2.27% Asian, and 0.20% Hispanic).
Results: Among the 39,908 participants, there were 9,128 AF cases. After adjustment for age, sex, and four principal components of ancestry, the PRS discriminated similarly across groups (area under curve [AUC] 0.65 in African, 0.68 in Asian, 0.66 in European and 0.67 in Hispanic). The odds ratio of the top 3% of polygenic scores compared to the bottom 97% was also similar across all groups with the largest associations in Europeans with an OR per SD (95% CIs) of 2.5 (2.2-2.8, p=1.3E-47). In African ancestry subjects the OR per SD was 1.7 (1.0-3.0, p=0.04), in Hispanic subjects the OR per was 2.9 (1.2-7.2, p=0.02) and 4.6 (1.3-16.7, p=0.02) in Asians.
Conclusions: Common genetic variation as defined by PRS is significantly associated with AF across multiple ancestries in the eMERGE Network. These findings suggest that despite the differing risk across ancestries, there are likely common genetic pathways that contribute to AF across ancestries. Future work will include evaluating integrative risk models for AF across diverse cohorts.

PrgmNr 3260 - Polygenic risk score for coronary artery disease worsens disparity by HIV status in clinical prediction of coronary artery calcium

View session detail

Author Block: **G. L. Wojcik**[1], C. I. Vergara[2], N. P. Arya[1], S. A. Haberlen[1], M. Budoff[3], M. J. Feinstein[4], J. W. Magnani[5], J. B. Margolick[1], S. Wolinsky[4], P. Duggal[1], W. S. Post[6]; [1]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD, [2]Johns Hopkins Univ, Baltimore, MD, [3]The Lundquist Inst., Univ. of California Los Angeles, Los Angeles, CA, [4]Northwestern Univ. Feinberg Sch. of Med., Chicago, IL, [5]Univ. of Pittsburgh Sch. of Med., Pittsburgh, PA, [6]Johns Hopkins Sch. of Med., Baltimore, MD
Disclosure Block: **G.L. Wojcik:** None.
The development and application of polygenic risk scores (PRS) for cardiovascular disease have moved us closer to the promise of precision medicine. However, the performance of such scores depends on the original study population used for their development being representative of patients with respect to both genetic and non-genetic factors. Their application can be problematic for subpopulations of patients not represented in the original study population, such as people living with human immunodeficiency virus (HIV) who have an increased risk for coronary artery disease (CAD) compared to those without HIV. To address this question, we assessed the predictive performance for coronary artery calcium (CAC) of a genome-wide PRS for CAD of 6.6 million variants (Khera 2018) in 731 HIV-infected and -uninfected men of European ancestry in the Multicenter AIDS Cohort Study (MACS), independently and in combination with the ACC/AHA Pooled Cohort Equation (PCE). CAC scores were calculated by the Agatston method from cardiac CT scans. Regression models were stratified by HIV serostatus and adjusted for the top 10 principal components to account for residual population substructure, age, CT scanning center, and cohort enrollment date (before or after 2001). The overall median age was 57 (interquartile range (IQR): 52-62) years, 66% with CAC > 0, median CAC 24 (IQR: 0-186), and 58% living with HIV (HIV+) of whom 89% were virologically suppressed (plasma HIV RNA 2 of 0.19 for HIV+ and 0.29 for HIV-) and for the dichotomous absence or presence of CAC (area-under-the-curve (AUC) of 0.75 for HIV+ and 0.81 for HIV-). This trend was consistent with the differential performance of the non-genetic PCE (adjusted R^2: 0.19 for HIV+ vs 0.23 for HIV-). When we added the PRS to the PCE estimates, the adjusted R^2 increased to 0.22 for HIV+ men and 0.28 for HIV- men, widening the performance gap between the two with the adjusted R^2 for HIV+ men being only 78% of that for HIV- men, compared with 82% for PCE alone. These results suggest that PRS for CAD may not adequately capture the gene x environment interactions which may differ between HIV+ and HIV- men when predicting CAC, resulting in reduced accuracy for HIV+ men relative to HIV- men when used alone or in conjunction with existing clinical risk scores. Future research must extend these questions to assessing risk performance for clinical CAD in people living with HIV in order to develop more accurate risk prediction models and ensure that PRS have clinical utility in all patient populations.

PrgmNr 3261 - The genetic determinants of body mass index & fasting glucose are mediators of grade 1 diastolic dysfunction development

View session detail

Author Block: **N. Vaitinadin**[1], M. Shi[1], C. M. Shaffer[1], E. Farber-Eger[2], B. D. Lowery[1], V. Agrawal[1], Q. S. Wells[2,1], J. D. Mosley[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Vanderbilt Univ., Nashville, TN
Disclosure Block: **N. Vaitinadin:** None.
Background Early (grade 1) cardiac diastolic dysfunction (G1DD) is a common functional change in the heart that increases the risk for heart failure (HF) with preserved ejection fraction and is treatable with aggressive risk factor modification. Type 2 diabetes (T2D), obesity, hypertension, and coronary heart disease are associated with increased incidence of diastolic dysfunction. The predisposing genetic drivers of G1DD are not defined. **Methods** We curated genotyped European Ancestry cases (n=668) and controls (n=1772) for G1DD from Vanderbiltâ□□s biobank, BioVU. GWAS, employing an additive genetic model was used to identify common variants associated with G1DD status. Common variant polygenic risk scores (PRS) for coronary heart disease, systolic blood pressure, T2D, and body mass index (BMI) were developed using data from prior GWAS for use in association studies and Mendelian randomization (MR) analyses. Logistic regression was used to test the associations between G1DD status and risk scores for each risk factor. Predictors with significant associations were further evaluated using 2-sample Mendelian randomization methods (the inverse-variance weighted [IVW], MR-Egger and median). Candidate mediator exposures were further characterized using multivariable mendelian randomization. **Results** There were no common SNPs associated with G1DD status at genome-wide significance by GWAS. A PRS for BMI was significantly associated with increased G1DD risk (OR = 1.20 for 1 SD increase in BMI, 95% CI = 1.08,1.32, p = 0.0003). The association was confirmed by IVW [OR=1.89 (95% CI=1.37, 2.61). Among the candidate mediators for BMI - HDL cholesterol, LDL cholesterol, triglycerides, fasting glucose (FG), and HbA1C - only FG was significantly associated with G1DD status by IVW (OR = 4.14 for 1 SD increase in FG, 95% CI = 1.55,11.02, p = 0.005). Multivariable mendelian randomization showed a modest attenuation of the BMI association [OR=1.84 (1.32, 2.57)] when adjusting for FG. **Conclusions** In sum, these data suggest that a genetic predisposition to elevated BMI increases the risk of G1DD. Part of this effect may be mediated through altered glucose homeostasis. BMI is an important modifiable risk factor for G1DD.

PrgmNr 3262 - The genetics of coronary artery calcification in individuals with type 2 diabetes

View session detail

Author Block: **N. R. Hasbani**[1], J. B. Meigs[2], S. Kwak[3], P. S. de Vries[4]; [1]Epidemiology, Human Genetics, and Environmental Sci.,The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Massachusetts Gen Hosp, Boston, MA, [3]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, [4]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Disclosure Block: **N.R. Hasbani:** None.

Coronary artery calcification (CAC) is a marker of atherosclerosis and is associated with increased risk of coronary heart disease (CHD) mortality, especially in individuals with type 2 diabetes (T2D). While numerous studies have identified genetic loci involved in CAC, CHD, and T2D, the shared genetic architecture between these highly associated traits is still being understood. We compared the effects of 207 genetic variants that were previously identified as associated with CHD and/or CAC in 2,971 individuals with T2D and 13,022 non-diabetic controls utilizing whole genome sequencing generated by the National Heart, Lung, and Blood Instituteâ□□s Trans-Omics for Precision Medicine (TOPMed) program. Participants were from four race/ethnic groups, including European American, African American, Hispanic/Latinx, and East Asian. CAC was first log transformed, then further transformed through inverse rank-based normalization of the residuals accounting for age and sex. Linear mixed models accounting for relatedness, implemented in GENESIS, were used to test for interaction between each variant and T2D status. Analyses were adjusted for age, sex, study and the first eleven principal components. The genetic main and interaction effects were assessed in a joint test using a two degree of freedom model to determine if a CHD variant was associated with CAC, then further evaluated to determine if these variants had a significantly different effect in T2D cases versus controls. Using Bonferroni corrected significance threshold of *P*-4 (0.05/207), we identified eighteen CHD variants associated with CAC according to the joint test, of which six had a statistically significant different effect in T2D cases and controls (rs6494488 near *RBPMS2,* rs7212798 in *BCAS3,* rs1321309 near *CDKN1A,* rs668948 near *APOB,* rs840616 near *CALCRL,* rs12897 near *FNDC3B*). The association of five of these variants was stronger in T2D cases than in controls, while for rs7212798 it was stronger in controls. While rs668948 has not previously been implicated in previous CAC GWAS, it lies nearest to *APOB,* which is a known driver of plaque development and subsequent atherosclerosis. Similarly, rs840616 lies near *CALCRL*, which is a part of a group of calcitonin receptors involved in the maintenance of calcium homeostasis. Overall, 85 of the CHD variants were nominally significant (*P*

PrgmNr 3263 - An exposome-wide association study (ExWAS) and comparison of clinical, environmental and polygenic risk scores for Type 2 Diabetes

View session detail

Author Block: **F. Akhtari**[1], D. T. Lloyd[1], J. S. House[1], D. C. Fargo[1], S. Schurman[2], C. S. Schmitt[1], J. E. Hall[1], A. A. Motsinger-Reif[1]; [1]NIEHS, Durham, NC, [2]Natl. Inst on Aging, NIH, Baltimore, MD
Disclosure Block: **F. Akhtari:** None.
Type 2 diabetes (T2D) is a complex disease influenced by genetic and environmental factors. Few studies have examined the effects of environmental exposures on T2D, especially in the context of genetic risk. We investigate the individual and cumulative effects of environmental and genetic factors on T2D, and report novel exposures associated with the disease. Analogous to a genome-wide association study, we conducted an exposome-wide association analysis (ExWAS) to assess the influence of diet, lifestyle, and environmental factors on T2D risk. We collected environmental and genetic data from the Personalized Environment and Genes Study (PEGS), a diverse North Carolina-based cohort, using three survey instruments: the Health and Exposure Survey (N = 9,414) for demographics, family medical history, lifestyle factors, and occupational exposures; the Internal Exposome Survey (N = 2,962) for exposures such as medications, physical activity, stress, sleep, and diet; and the External Exposome Survey (N = 3,519) for exposures such as chemical and environmental exposures at work and home. We performed single-exposure association analysis for 680 environmental factors with self-reported T2D status using logistic regression models with appropriate covariate adjustment. After multiple testing correction, we identified 27 significant associations. Notably, we identified novel associations of exposure to asbestos (OR = 1.44, CI = 1.10-1.88), textiles (OR = 1.39, CI = 1.05-1.83), and coal dust (OR = 1.60, CI = 1.05-2.43) with T2D. To compare cumulative risk due to genetic and non-genetic factors, we computed clinical risk scores (CRS), environmental risk scores (ERS), and polygenic risk scores (PRS) for a subset of participants (N = 3,236) for whom exposure data and whole genome sequencing was available. We calculated CRS using self-reported BMI, pre-diabetes, hypertension, and high cholesterol. We computed ERS using 16 variables selected by LASSO from the significant ExWAS results. We developed a trans-ethnic PRS for T2D with LDPred2 using UK Biobank data (N = 440,757; 122,359 SNPs) and then calculated the PRS for PEGS participants. Using logistic regression models adjusted for age, gender, and the first ten principal components, we found that all three risk scores were significantly associated with T2D, with odds ratios for CRS (3.85, CI = 3.38-4.42), ERS (2.90, CI = 2.49-3.40) and PRS (1.29, CI = 1.11-1.52). Notably, ERS had a stronger association and higher correlation with T2D compared to the more frequently reported PRS. Future work should replicate this finding and assess the predictive performance of PRS. In ongoing work, we are evaluating PRSxE interactions.

PrgmNr 3264 - Assessment of Polygenic Risk Score for Predicting Type 2 Diabetes Risk in Asian Indians: Results of the Asian Indian Diabetic Heart Study/Sikh Diabetes Study

View session detail

Author Block: **D. K. Sanghera**[1], S. Goyal[1], J. R. Singh[2], N. K. Mehra[3], G. S. Wander[4], S. Ralhan[4]; [1]Univ Oklahoma HSC, Oklahoma City, OK, [2]Central Univ. of Punjab, Bathinda, India, [3]All India Inst. of Med. Sci. and Res., New Delhi, India, [4]Hero DMC Heart Inst., Ludhiana, India
Disclosure Block: **D.K. Sanghera:** None.
Genome-wide polygenic risk score (PRS) is considered a good predictor of disease risk, however, its utility as an independent risk predictor in complex diseases such as type 2 diabetes (T2D) remains inconclusive. Epidemiological studies reported that people from the Indian subcontinent suffer from a greater risk of cardiovascular disease-associated mortality than other global populations. Overwhelming data support a strong influence of genetic factors and their interaction with lifestyle and environmental factors. Populations from the Indian subcontinent are underrepresented in genetic studies despite Asian Indians contribute more than 25% of the world population and have up to 6 times more risk for T2D compared to European whites. Here we have evaluated the additive value of multilocus PRS in T2D risk-prediction in comparison to the Clinical Risk Score (CRS, modified from the Joint British Society (JBS) risk score) using 4,588 individuals (2,566 cases and 2,022 controls) from the Asian Indian Diabetes Heart Study/Sikh Diabetes Study (AIDHS/SDS). The PRS for T2D was constructed using 63 genetic variants associated with T2D in 1,613 individuals in the training set and validated in 2,975 individuals in the test set. Logistic regression analyses were performed using age, gender, BMI, and five principal components as covariates to determine the risk prediction for T2D using PRS and CRS. Prediction accuracy of PRS and CRS was assessed by measuring the area under the receiver operating characteristic (ROC) curve (AUC). PRS for T2D yielded an adjusted odds ratio (OR)1.86 95% CI (1.73-2.00), $p=2.39x10^{-62}$ with AUC of 0.78 (95% CI 0.75-0.80) in the training set, and OR1.51 95% CI (1.43-1.59), $p=5.64x10^{-55}$ with AUC of 0.76; 95% CI (0.75-0.78) for the test set. In comparison to the PRS, the composite analysis of CRS revealed a higher OR and increased AUC in both training and the test sets, respectively [OR 2.63 95%CI (2.36-2.92), $p=1.58x10^{-69}$ and AUC of 0.80 (95% CI 0.78-0.82)] and [OR 2.7195% CI (2.50-2.95), $p=1.66x10^{-123}$, and an AUC of 0.84; 95% CI (0.83-0.85)]. In summary, though the PRS was an independent predictor of the T2D risk in this study, the CRS has still outperformed the PRS. Nevertheless, as more risk loci with population-specific effects are being identified by sequencing, combining the PRS approach with clinical/traditional risk factors would enhance predictive value to CRS to identify and stratify individuals at higher risk of developing T2D.

PrgmNr 3265 - Do methods of inclusion and presentation of non-European ancestry data in GWAS allow for novel discoveries?

View session detail

Author Block: **A. R. Bentley**[1], C. N. Rotimi[2]; [1]Natl Human Genome Res. Inst., Bethesda, MD, [2]NIH, Bethesda, MD
Disclosure Block: **A.R. Bentley:** None.
As is well appreciated, studies of European ancestry populations (EUR) dominate the GWAS literature. The call for greater inclusion of non-EUR studies has been partly motivated by the potential for novel discoveries. As more non-EUR populations are being included in GWAS publications, it is worth evaluating how these data are being incorporated, and if the study designs and presentation allow for the scientific advances that motivate their inclusion. We investigated GWAS studies published from 2015-2021 of type 2 diabetes (T2D) and related traits (retrieved from the GWAS catalog 5/13/2021) and sought to characterize them in terms of whether the study design and presentation allow for the discovery of novel associations in the non-EUR data.
Of the 102 publications that met our study criteria, 44% focused exclusively on EUR, both in discovery and replication, and are not discussed further. Of the 57 publications that included non-EUR studies in some way, the degree to which the inclusion facilitated novel findings varied. Three studies only included non-EUR studies to evaluate generalizability of EUR GWAS findings, thus limiting non-EUR associations to those already discoverable in EUR. Publications focused on non-EUR populations (n=35) were dominated by those of East Asian ancestry (n=23), but also included studies of Hispanics (n=4), Native Americans (n=2), Arabic individuals (n=2), Africans (n=2), African Americans (n=1), and Aboriginal Australians (n=1). There were 19 publications that conducted discovery GWAS in multiple ancestry groups. Eight of these only presented trans-ancestry meta-analyses, which do not allow for future ancestry-specific meta-analyses and may not adequately capture novel discoveries in each ancestry in the presence of heterogeneity of effects or linkage disequilibrium differences. In total, 19% of the studies including non-EUR populations were not optimized in design and/or presentation for facilitating genomic discoveries in these populations.
While studies have characterized the number and sample sizes of included studies of diverse ancestries, it is not only inclusion of non-EUR ancestry studies that matters. To maximize the potential for novel findings, it is also important how these cohortsâ data are analyzed and interpreted. To take advantage of the genetic diversity in non-EUR populations, it is key to conduct and present discovery studies in these populations separately. It is preferable to accompany trans-ancestry meta-analyses with ancestry-specific analyses and presentations.

PrgmNr 3266 - Expanded Genetic Clustering of Type 2 Diabetes Loci Using a High-throughput Pipeline Reveals Mechanistic Pathways of Metabolic Diseases

View session detail

Author Block: **H. Kim**[1,2], K. Westerman[1,2], J. Chiou[3], J. B. Cole[2,1,4,5], T. D. Majarian[2], M. von Grotthuss[6], J. M. Mercader[2,1,4], S. Kwak[7], J. Kim[8], J. C. Florez[1,2,4], K. Gaulton[9], A. K. Manning[1,2,4], M. S. Udler[10,1,4]; [1]Massachusetts Gen. Hosp., Boston, MA, [2]The Broad Inst. of MIT and Harvard, Cambridge, MA, [3]UC San Diego, La Jolla, CA, [4]Harvard Med. Sch., Boston, MA, [5]Boston Children's Hosp., Boston, MA, [6]Takeda Pharmaceuticals, Cambridge, MA, [7]Seoul Natl. Univ. Hosp., Seoul, Korea, Republic of, [8]GlaxoSmithKline, Cambridge, MA, [9]UC San Diego, San Diego, CA, [10]Broad Inst., Cambridge, MA
Disclosure Block: **H. Kim:** None.
Complex diseases such as type 2 diabetes (T2D) are highly polygenic and influenced by multiple biological pathways. Rapid expansion in the number of T2D loci can be leveraged to identify such pathways, which may facilitate improved patient management.
We developed a high-throughput pipeline to enable clustering of T2D loci based on variant-trait associations. Our pipeline extracted summary statistics from genome-wide association studies (GWAS) for T2D and related trait to generate a matrix of 336 variants x 64 trait associations and applied Bayesian Non-negative Factorization (bNMF) to identify genetic components of T2D.
We identified ten genetic clusters of T2D, which included five from our published prior analysis of 94 T2D loci. Four of the ten clusters related to mechanisms of insulin deficiency, five to insulin resistance, and one had an unclear mechanism. Novel clusters identified in this analysis related to beta-cell dysfunction, pronounced insulin secretion, and circulating levels of alkaline phosphatase, lipoprotein A, and sex hormone binding globulin.
The T2D genetic clusters displayed tissue-specific epigenomic enrichment, particularly in pancreatic islets, liver, and adipose tissue. Two of the clusters relating to insulin deficiency were differentially enriched in functional and stressed pancreatic beta-cell states. Additionally, cluster-specific polygenic scores were associated with distinct clinical outcomes across GWAS and confirmed in participants in the Mass General Brigham Biobank. Multiple observed T2D genetic pathways were shared across genetic clusters of coronary artery disease and chronic kidney disease.
Our approach stratifies T2D loci into physiologically meaningful clusters with distinct tissue specificity and association with metabolic conditions. The pipeline allows for efficient updating and refining of clusters as additional GWAS datasets become available, and can be readily applied to other conditions. This method supports translation of GWAS findings into a more granular understanding of disease mechanisms, with a view toward precision medicine.

PrgmNr 3267 - Genetic Regulation of Obesity Explains Cardiovascular Sex Differences in Polycystic Ovary Syndrome for Predisposed Individuals

View session detail

Author Block: **K. Actkins**[1], L. K. Davis[2]; [1]Meharry Med. Coll., Nashville, TN, [2]Vanderbilt Univ Med Ctr., Nashville, TN
Disclosure Block: **K. Actkins:** None.
Introduction: Polycystic ovary syndrome (PCOS) is the most common endocrine disorder in reproductive age women with a complex polygenic architecture largely defined by metabolic dysregulation. Despite only being diagnosable in females, males with a family history of PCOS can also exhibit a poorer cardiometabolic profile that can be detected as early as infancy. In our previous phenome-wide association study (PheWAS), we found that males with a higher polygenic risk score (PRS) for PCOS ($PCOS_{PRS}$) were more likely to develop cardiovascular diseases like hypertension compared to females who had higher odds of developing type 2 diabetes (T2D). Therefore, in this study, we aimed to further elucidate the role of sex in the relationship between PCOS and its comorbidities by evaluating bidirectional genetic pathways. **Methods:** To do this, we first calculated PRS for BMI, T2D, systolic and diastolic blood pressure, and coronary artery disease (CAD) (i.e., the phenome-wide significant phenotypes identified in the prior PheWAS). Each PRS was then fitted against PCOS diagnosis in a logistic regression model adjusted for median age, genetic ancestry, and clinically measured body mass index (which captured both genetic and environmental BMI variance). Next, to examine the independent effects of environmental BMI, genetically regulated BMI variance was regressed out of clinical BMI measurements. Mediation analyses were then conducted in which BMI or residual BMI (i.e., with genetic variance removed) were coded as the mediator, $PCOS_{PRS}$ was coded as the exposure variable, and the above dichotomized clinical diagnoses as outcomes. **Results:** Only $T2D_{PRS}$ was significantly associated with PCOS in females (OR = 1.16[1.05-1.28], p=3.69e-03). This association was attenuated when including BMI as a covariate (OR = 1.07[0.97-1.19], p=0.20). However, the association reappeared when residual BMI was covaried (OR = 1.18[1.06-1.30], p=2.07e-03). Unsurprisingly, clinically measured BMI variance significantly mediated the effect of $PCOS_{PRS}$ on cardiometabolic outcomes in both sexes, but residual BMI alone did not mediate the effect of $PCOS_{PRS}$ on T2D (p=0.78) nor $PCOS_{PRS}$ on hypertension (p=0.82) in males. **Conclusions:** Our findings show that genetically predicted BMI has a greater relative impact than residual BMI on PCOS and its cardiometabolic comorbidities. Moreover, the genetic architecture of PCOS results in distinct metabolic sex differences. It is possible that comorbid conditions (e.g., T2D, CAD) may increase risk for PCOS through shared metabolic genetic pathways.

PrgmNr 3268 - Genotype-Phenotype Analysis in African Americans with Inflammatory Bowel Disease

View session detail

Author Block: **A. Stiemke**[1], L. W. Datta[2], M. G. Lazarev[2], Inflammatory Bowel Disease Genetics Consortium, D. P. B. McGovern[3], S. R. Brant[4,2], C. L. Simpson[1]; [1]Genetics, Genomics and Informatics, Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [2]Meyerhoff Inflammatory Bowel Disease Ctr., Dept. of Med., Sch. of Med. and Dept. of Epidemiology, Sch. of Publ. Hlth., Johns Hopkins Univ., Baltimore, MD, [3]F. Widjaja Fndn. Inflammatory Bowel and Immunobiology Res. Inst., Cedars Sinai Med. Ctr., Los Angeles, CA, [4]Rutgers Robert Wood Johnson Med. Sch., New Brunswick, NJ
Disclosure Block: **A. Stiemke:** None.
Inflammatory bowel disease (IBD) is an immune-mediated chronic intestinal disorder that is typically divided into two distinct types; ulcerative colitis (UC) and Crohn's disease (CD). Although each respective disease is classified based on clinical presentation, each diseaseâ□□s presentation is highly heterogenous. Here we aimed to better discern the heterogeneity within disease types by characterizing the many endophenotypes within UC and CD rather than their initial classification. A total of 1135 African American patients, 772 with CD, 323 with UC, and 41 with inflammatory bowel disease unclassified (IBDU), were included in this study. Genomic DNA data on the HumanOmni2.5 microarray was obtained from all patients and mapped to GRCh37/hg19. Extensive phenotyping was conducted on all patients, which allowed for the creation of 14 direct and derived endophenotypes, including perianal disease, smoking status, and age of onset. Analysis was restricted to the area Â±500kb around 15 previously identified candidate single nucleotide polymorphisms previously documented in IBD GWAS. Statistically significant results have been found for a number of stratified endophenotype analyses including disease severity and disease behavior. One such result for this stratification (*P*-value 2.64e-6, OR 0.21) is approximately 380kb proximal to NOD2 in ADCY7. Additional analyses are ongoing. Utilizing endophenotypes to differentiate various disease presentations will allow for additional genetic markers to be discovered. As additional genetic risk loci are identified, more robust screening will improve patient outcomes and overall quality-of-life. If the disease can be caught in its early stages, treatment options can be taken to ensure the best patient outcome.

PrgmNr 3269 - Identification of hepatocellular gene expression phenotypes influencing fatty liver disease in Hispanics

View session detail

Author Block: **S. Kumar**[1], J. E. Curran[2], E. De Leon[1], J. C. Granados[1], A. C. Leandro[2], M. Leandro[2], J. M. Peralta[2], S. Williams-Blangero[2,1], J. Blangero[2]; [1]Dept. of Human Genetics and South Texas Diabetes and Obesity Inst., Univ. of Texas Rio Grande Valley Sch. of Med., McAllen, TX, [2]Dept. of Human Genetics and South Texas Diabetes and Obesity Inst., Univ. of Texas Rio Grande Valley Sch. of Med., Brownsville, TX

Disclosure Block: **S. Kumar:** None.

Fatty liver disease (FLD) encompasses conditions that are associated with excess fat deposits in the liver (hepatic steatosis) and is the common cause of chronic liver disease worldwide. Non-alcoholic FLD (NAFLD) has been associated with a wide range of risk factors including diabetes, obesity, dyslipidemia, and metabolic syndrome and disproportionately affects the medically underserved Hispanic minority in the United States. NAFLD risk is likely due to a complex interaction of genetic and environmental factors that are still largely unidentified. We used an induced pluripotent stem cell (iPSC) based disease modeling approach to identify novel gene expression phenotypes influencing the disease to better understand disease genetics and develop targeted treatments. Well characterized, iPSC differentiated functional hepatocytes (HEPs) generated from six participants of our San Antonio Mexican American family study (SAMAFS) were analyzed for cellular lipid accumulation in *in-vitro* culture and by genome wide mRNA sequencing. The quantitative measures of cellular lipids, obtained by staining neutral lipids with BODIPY lipid probe and high content screening analysis of the iPSC generated HEPs, showed a high correlation (r^2 = 70%) with individualâ□□s *in-vivo* liver fat measures obtained by magnetic resonance imaging (MRI). Genome wide differential gene expression analysis between HEPs generated from individuals with high liver fat (MRI liver fat â□¥ 8%; n = 4) and normal/low liver fat (MRI liver fat â□¤ 6%; n = 2) identified 104 genes that were significantly differentially expressed (moderated t statistics p-value â□¤ 0.05, fold change (FC) absolute â□¥ 2.0). Functional annotation and gene network analyses performed using Ingenuity Pathway Analysis platform showed significant enrichment of differentially expressed genes in lipid metabolism (p- value range $9.11x10^{-5}$ to $7.84x10^{-3}$), small molecule biochemistry (p- value range $9.11x10^{-5}$ to $7.89x10^{-3}$) and cell-to-cell signaling (p- value range $2.43x10^{-4}$ to $1.17x10^{-2}$) cellular and molecular functions. Expression of several genes involved in lipid oxidation (*HAO2*, *RGS16* and *MB*), and lipid and sphingolipid homeostasis (*ORM1, ORM2, PON3,* and *ABCG1*) were found significantly altered in HEPs of high liver fat individuals. The expression of *ABCG1* which plays an important role in deposition and mobilization of cholesterol and overall lipid homeostasis in the liver was found to be significantly downregulated. An *in-vitro* fatty acid challenge performed on HEPs showed significant (FC â□¥ 2.0) upregulation of *ABCG1* in normal individuals, however nominal upregulation was observed in high liver fat individuals.

PrgmNr 3270 - Investigating the Causal Role of Reduced Vitamin D Levels With Type 2 Diabetes Risk in South Asians and Europeans

View session detail

Author Block: **S. GOYAL**[1], C. A. Bejar[1], S. Afzal[2], M. Mangino[3], A. Zhou[4], P. J. van der Most[5], Y. Bao[6], V. Gupta[7], M. C. Smart[6], G. K. Walia[8], N. Verweij[9], C. Power[10], D. Prahbakaran[8], J. R. Singh[11], N. K. Mehra[12], G. S. Wander[13], S. Ralhan[13], S. Kinra[14], M. Kumari[6], M. H. de Borst[9], E. HyppÃ¶nen[4], T. D. Spector[15], B. G. Nordestgaard[16], P. R. Blackett[17], D. K. Sanghera[17]; [1]Oklahoma Univ. of Hlth.Sci. Ctr., Oklahoma, OK, [2]Herlev and Gentofte Hosp., Copenhagen Univ. Hosp., Herlev, Denmark, [3]Kings Coll. London, London, United Kingdom, [4]Australian Ctr. for Precision Hlth., Univ. of South Australia Cancer Res. Inst., Adelaide, Australia, [5]Dept. of Epidemiology, Univ. of Groningen, Univ. Med. Ctr. Groningen, Groningen, Netherlands, [6]Univ. of Essex, Colchester, United Kingdom, [7]Dept. of Anthropology, Univ. of Delhi, New Delhi, India, [8]Publ. Hlth.Fndn. of India, Gurgaon, India, [9]Dept. of Internal Med., Div. of Nephrology, Univ. of Groningen, Univ. Med. Ctr. Groningen, Groningen, Netherlands, [10]Population, Policy and Practice, Inst. of Child Hlth., Univ. Coll. London, London, United Kingdom, [11]Central Univ. of Punjab, Bathinda, India, [12]All India Inst. of Med. Sci. and Res., New Delhi, India, [13]Hero DMC Heart Inst., Ludhiana, India, [14]London Sch. of Hygiene and Tropical Med., London, United Kingdom, [15]Dept. of Twin Res. and Genetic Epidemiology, Kings Coll. London, London, United Kingdom, [16]Dept. of Clinical Biochemistry, Herlev and Gentofte Hosp., Copenhagen Univ. Hosp., Herlev, Denmark, [17]Section of Pediatric Endocrinology, Coll. of Med., Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma, OK
Disclosure Block: **S. Goyal:** None.

Multiple observational studies have reported an inverse relationship between 25-hydroxyvitamin D concentrations (25(OH)D) and type 2 diabetes (T2D). However, the results of short- and long-term interventional trials concerning the relationship between 25(OH)D and T2D risk have been inconsistent. To evaluate the causal role of reduced blood 25(OH)D in T2D, here we have performed a bidirectional Mendelian randomization study using 59,890 individuals (5,862 T2D cases and 54,028 controls) from European and Asian Indian ancestries. We used six known SNPs including 3 T2D SNPs and 3 vitamin D pathway SNPs as a genetic instrument to evaluate causality and direction of the association between T2D and circulating 25(OH)D concentration. Results of the combined meta-analysis of eight participating studies showed that a composite score of three T2D SNPs would significantly increase T2D risk by odds ratio of 1.24, $p=1.82x10^{-32}$; Z score 11.86 which however had no significant association with 25(OH)D status (Beta -0.02nmol/L Â± SE 0.01nmol/L; p=0.83; Z score -0.21). Likewise, the genetically instrumented composite score of 25(OH)D lowering alleles significantly decreased 25(OH)D concentrations (-2.1nmol/L Â± SE 0.1nmol/L, $p=7.92x10^{-78}$; Z score -18.68) but was not associated with increased risk for T2D (odds ratio 1.002, p=0.12; Z score 1.54). However, using 25(OH)D synthesis SNP (DHCR7; rs12785878) as an individual genetic instrument, a per allele reduction of 25(OH)D concentration (-4.2nmol/L Â± SE 0.3nmol/L) was predicted to increase T2D risk by 5%, p= 0.004; Z score 2.84. This effect, however, was not seen in other 25(OH)D SNPs (GC rs2282679, CYP2R1 rs12794714) when used as an individual instrument. Our new data on this bidirectional Mendelian randomization study suggests that genetically instrumented T2D risk does not cause changes in 25(OH)D levels. However, genetically regulated 25(OH)D deficiency due to vitamin D synthesis gene (DHCR7) may influence the risk of T2D.

PrgmNr 3271 - Pediatrics and proteins: Molecular discovery in the Hearts & Parks study of children with obesity

View session detail

Author Block: **N. A. Bihlmeyer**[1], J. McCann[2], J. F. Rawls[2], A. C. Skinner[3], S. C. Armstrong[4], S. H. Shah[5]; [1]Duke Univ., Durham, NC, [2]Duke Molecular Genetics and Microbiol., Durham, NC, [3]Duke Clinical Res. Inst., Durham, NC, [4]Duke Ctr. for Childhood Obesity Res., Durham, NC, [5]Duke Univ, Durham, NC
Disclosure Block: **N.A. Bihlmeyer:** None.
Obesity is occurring at earlier ages in contemporary society. Mirroring adults, one third of US children are overweight or obese. This portends a worsening epidemic of the downstream consequences of obesity including cardiovascular disease. Unfortunately, similar to studies in adults, there is heterogeneity with regards to weight and metabolic response to obesity interventions in children which is poorly characterized with clinical factors. Thus, we sought to use proteomic profiling to elucidate biology underlying this heterogeneity and identify biomarkers with potential use for personalized weight loss interventions. Proteomic profiling of 768 proteins using the Olink proximal extension platform was performed in frozen baseline serum samples from two behavioral weight loss intervention cohorts: (1) Hearts & Parks (H&P; N=116), a lifestyle-based weight loss study of children (5-12 yo); (2) Pediatric Obesity Microbiome & Metabolism Study (POMMS; N=184) of adolescents (10-18 yo). Outcomes included: baseline and percent change in Homeostatic Model Assessment of Insulin Resistance (%â□□HOMA-IR); and obese/lean status and change in BMI P95. Linear/logistic models were used to associate proteins to outcomes. The mean baseline HOMA-IR was: H&P = 3.2Â±3.3; POMMS = 3.8Â±4.5; and the mean %â□□HOMA-IR was: H&P = 0.2Â±0.7; POMMS = 0.7Â±1.1. The mean change BMI P95 was H&P = -1.0Â±6.7; POMMS = 1.0Â±7.0. Proteomic profiling identified 328 proteins associated with obese vs lean, two with â□□BMI P95, 66 with baseline HOMA-IR, and none with %â□□HOMA-IR, all after FDR correction for multiple testing. IGFBP-1 and IGFBP-2 (both correlated with r2=0.64) were the most significantly associated with obese/lean status (FDR p

PrgmNr 3272 - Phenome-wide investigation of health outcomes associated with the genetic correlates of 25 hydroxyvitamin D concentration

View session detail

Author Block: **H. Kresge**[1], S. Goleva[1], J. A. Revez[2], N. Wray[3], J. J. McGrath[2,4,5], L. K. Davis[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Univ. of Queensland, Brisbane, Australia, [3]The Univ. of Queensland, Brisbane, Australia, [4]Queensland Ctr. for Mental Hlth.Res., The Park Ctr. for Mental Hlth., Wacol, Australia, [5]Aarhus Univ., Aarhus, Denmark
Disclosure Block: **H. Kresge:** None.
Vitamin D deficiency, a prevalent condition that is commonly associated with poor general health, may place individuals at increased risk for subsequent disorders. Previous research demonstrated links between genetic correlates of vitamin D concentration and dyslipidemia, coronary artery disease, and type 2 diabetes. This study aims to explore associations between common genetic variants associated with 25 hydroxyvitamin D (25OHD) concentration and a comprehensive range of general clinical phenotypes and laboratory tests.

25OHD-related polygenic risk scores (PRS_25OHD) were generated for Vanderbilt University Medical Center Biobank (BioVU) participants of European ancestry (EA, n=72,741) and African ancestry (AA, n=15,283). 25OHD_PRS were used to perform a phenome-wide association study (PheWAS) and laboratory test-wide association study (LabWAS).

In EA PheWAS models, PRS_25OHD was negatively associated with clinical vitamin D deficiency (OR=0.84, 95% CI=0.81-0.87, p=2E-32), hypercholesterolemia (OR=0.91, 95% CI=0.88-0.94, p=3E-11), disorders of lipid metabolism (OR=0.94, 95% CI=0.92-0.96, p=2E-10), hyperlipidemia (OR=0.94, 95% CI=0.92-0.96, p=2E-10), hyperglyceridemia (OR=0.79, 95% CI=0.70-0.88, p=3E-07), type 2 diabetes (OR=0.94, 95% CI=0.92-0.96, p=3E-08), and obesity (OR=0.95, 95% CI=0.92-0.98, p=1E-06). In AA PheWAS models, PRS_25OHD was not associated with any phenotypes. In EA LabWAS models, PRS_25OHD was positively associated with observed concentrations of 25OHD ($\hat{I}^2$=0.18, SE=0.01, p=4E-82) and 1,25 hydroxyvitamin D, the hormonally active form of vitamin D ($\hat{I}^2$=0.15, SE=0.01, p=2E-120). PRS_25OHD was negatively associated with levels of triglycerides ($\hat{I}^2$=-0.06, SE=0.01, p=9E-27), cholesterol ($\hat{I}^2$=-0.04, SE=0.01, p=5E-16), and cholesterol in low density lipoproteins ($\hat{I}^2$=-0.03, SE=0.01, p=1E-06). In AA LabWAS models, PRS_25OHD was positively associated with observed concentrations of 1,25 hydroxyvitamin D ($\hat{I}^2$=0.07, SE=0.02, p=2E-5).

This study confirms that genetic instruments related to vitamin D status can predict a range of medical conditions and laboratory measures within electronic health records. PRS_25OHD was significantly associated with clinically defined vitamin D deficiency and laboratory measures of both 25OHD and the hormonally active form of vitamin D. Furthermore, lower genetically predicted vitamin D levels were associated with increased risk of lipid-related phenotypes, obesity, and diabetes. Future work is needed to elucidate mechanisms underlying associations reported in this study to understand how vitamin D deficiency may contribute to subsequent clinical manifestations.

PrgmNr 3274 - Rare variants associated with serum creatinine in 350k UK Biobank whole exome sequences

View session detail

Author Block: **M. M. Parker**[1], A. Deaton[2], L. D. Ward[1], R. Hoffing[2], C. Willis[1], M. Plekan[2], A. Flynn-Carroll[2], G. Hinkle[2], P. Nioi[1]; [1]Alnylam, Cambridge, MA, [2]Alnylam Pharmaceuticals, Cambridge, MA
Disclosure Block: **M.M. Parker:** Salary/Employment; Alnylam Pharmaceuticals.
Background High serum creatinine is an indicator of poor kidney function and is strongly negatively correlated with estimated glomerular filtration rate (eGFR). **Methods** We aggregated variants by gene and tested if rare (MAF P of $3.4x10^{-6}$ was considered statistically significant. Genes associated with creatinine were tested for association with a diagnosis of chronic kidney disease (CKD), with a *P* Results We find 11 significant associations between aggregated pLOF variants and serum creatinine. The strongest associations of pLOF variants with serum creatinine were in *PKD1* (OR CKD = 6.1) and *PKD2* (OR CKD =2.3), which are known causes of polycystic kidney disease. Of 638 carriers of pLOF variants in *PKD1*, 102 (16%) were in were diagnosed with CKD and of 125 carriers of pLOF variants in *PKD2*, 31 (30%) were diagnosed with CKD. Other genes associated with both lower creatinine and increased risk of CKD included *SLC22A2* (P creatinine = $1.7x10^{-38}$, P CKD = $1.0x10^{-7}$, OR CKD=1.7), *SLC34A3* (P creatinine = $4.0x10^{-14}$, P CKD = $1.9x10^{-3}$, OR CKD=1.7) and *OBSCN* (P creatinine = $2.3x10^{-9}$, P CKD = $4.1x10^{-3}$, OR CKD=1.2). Genetic variants in or around *SLC22A2* and *OBSCN* have been previously associated with eGFR in GWAS. Our rare variant associations suggest that loss of function in these genes increases risk of CKD, providing a potential mechanism to explain the GWAS associations. Similarly, genetic variants in *SLC34A3* have been previously associated with kidney stones and nephrocalcinosis, which our data suggests may be explained by loss of function of *SLC34A3*. **Conclusions** Tests between rare pLOF variants in 14,863 genes and serum creatinine revealed 11 significant associations, 5 of which also showed an association with increased risk of CKD diagnosis. Our results help explain previously identified GWAS associations and will aid the advancement of precision medicines like RNAi therapeutics.

PrgmNr 3275 - The effect of obesity-related traits on COVID-19 severe respiratory symptoms and hospitalization and its mediation by socioeconomic status: a multivariable Mendelian randomization study

View session detail

Author Block: **B. Cabrera Mendoza**[1,2], F. Wendt[1,2], G. Pathak[2,1], F. de Angelis[1,2], A. De Lillo[1], D. Koller[1,2], R. Polimanti[1,2]; [1]Yale Univ., New Haven, CT, [2]VA CT Hlth.Care System, West Haven, CT
Disclosure Block: **B. Cabrera Mendoza:** None.
Obesity has been associated with a higher susceptibility to coronavirus disease 2019 (COVID-19), particularly with its more severe clinical manifestations. However, this association can be affected by many correlates of these traits. Due to its large impact on human health, socioeconomic status (SES) could influence at least partially the association between obesity and COVID-19 severity. To estimate the independent effect of traits related to body size and SES on the clinical manifestations of COVID-19, we conducted a Mendelian randomization (MR) study analyzing the effect of obesity-related anthropometric traits on COVID-19 outcomes. We evaluated the effects of body mass index (BMI), waist circumference (WC), hip circumference, (HIP) and waist-hip ratio (WHR) studied in up to 234,069 participants from the Genetic Investigation of ANthropometric Traits (GIANT) consortium with respect to three COVID-19 outcomes: severe respiratory COVID-19 (5,101 cases vs. 1,383,241 controls), hospitalized COVID-19 (9,986 cases vs. 1,877,672 controls), and COVID-19 infection (38,984 cases vs. 1,644,784 controls) obtained from the COVID-19 Host Genetics Initiative (HGI). Finally, to investigate the effect of SES, we analyzed genetic data related to self-reported household income (HI) information from 286,301 UK Biobank (UKB) participants. We found that BMI and WC were associated with severe respiratory COVID-19 (BMI: ORâ□□=â□□1.68 pâ□□=â□□0.0004; WC: ORâ□□=â□□1.72, pâ□□=â□□0.0072) and hospitalized COVID-19 (BMI: OR =â□□1.62, pâ□□=â□□1.35e-06; WC: ORâ□□=â□□1.62, pâ□□=â□□0.0001). Also, HIP showed to influence hospitalized COVID-19 (ORâ□□=â□□1.31, pâ□□=â□□0.012) and COVID-19 infection (ORâ□□=â□□1.18, pâ□□=â□□0.0016). Conversely, HI was associated with reduced severe respiratory COVID-19 (ORâ□□=â□□0.57, pâ□□=â□□0.011) and hospitalized COVID-19 (ORâ□□=â□□0.71, pâ□□=â□□0.045). Testing these effects in multivariable MR models, we observed that the effect of obesity-related anthropometric traits on COVID-19 outcomes is not independent of the SES effect assessed as HI.. In summary, our findings indicate that low SES is a contributor to the observed association between body size and COVID-19 outcomes. Thus, the association of obesity with COVID-19 outcomes may be due to the conditions related to low SES rather than pathogenic mechanisms linked to obesity. This result has major public health implications because it supports that preventive strategies targeting body size and composition to reduce COVID-19 morbidity and mortality may not be effective if they are not considered in the context of SES.

PrgmNr 3276 - Whole exome sequencing analyses of BMI and obesity in a multi-ancestry cohort of > 300,000 individuals

View session detail

Author Block: **N. Chami**[1], J. A. Brody[2], C-T. Liu[3], K. E. North[4], A. Justice[5], R. Loos[6], on behalf of the TOPMed Anthropometry Working Group; [1]The Charles Bronfman Inst. for Personalized Med., Mount Sinai, New York, NY, [2]Univ of Washington, Seattle, WA, [3]Boston Univ. SPH, Boston, MA, [4]Univ North Carolina, Chapel Hill, NC, [5]Geisinger, Danville, PA, [6]The Icahn Sch. of Med. at Mount Sinai, New York, NY

Disclosure Block: **N. Chami:** None.

Background: Obesity is a major risk factor for T2D and CVD. GWA studies identified hundreds of common variants associated with BMI and obesity. In contrast, exome-chip analyses by the GIANT consortium in 718,734 individuals suggest that large-scale sequencing data is required to identify truly rare variants. Here, we aimed to identify rare (MAF Methods: We performed ancestry-specific single variant (SV) association analysis of BMI and obesity within Europeans (N=234,719), Africans (N=31,452), Hispanics (n=21,687), Asians (n=8,437), and other ancestry groups (N=3,375). We restricted analyses to protein truncating & missense variants with MAF10. We analyzed 1.2 million variants and used a Bonferroni-corrected P-value threshold of $4x10^{-8}$ for significance. Analyses were performed using GENESIS and regenie. Summary statistics were meta-analyzed across datasets using fixed effects. Gene-based analyses is underway.**Results:** None of the variants reached the Bonferroni-corrected threshold for our SV analysis. However, among our most significant results, we identify rare variants in thirteen biologically relevant genes. For example, p.Arg35Gln in *WDTC1* (MAF_{AFR}=0.5%) is associated with obesity (OR=2.4, $P=4x10^{-7}$) in the African-ancestry population, whereas this variant was much rarer in other ancestries and monomorphic in Europeans. As WDTC1 plays a role in adipogenesis and lipid accumulation in mouse models, the gene may present as a new candidate for human obesity. Also p.Gly85Glu in *MMP20* (MAF_{EUR}: 0.08%) was associated with obesity (OR=1.5, $P=6x10^{-7}$). MMP20 belongs to the matrix metalloproteinase family, which includes several other proteins implicated in obesity. Additionally, we observe many novel rare variants (P-6) in genes (*TNRC18, ZNF687, PARG1, USH2A and GEM*) in which common variants have been associated with BMI, extreme obesity or height. We note that the well-known monogenic obesity mutation, p.Tyr35Ter in *MC4R* (MAF=0.01%, $P=8.5x10^{-6}$; $Beta_{BMI}$=0.74) and the previously reported p.Gly245Arg in *SPARC* (MAF 0.04%; $P=4.9x10^{-6}$; $Beta_{BMI}$=0.07) did not reach our corrected P-value threshold.**Conclusions:** We observed thirteen rare variant associations that failed to reach the exome-wide Bonferroni-corrected P-value but that lie in genes that have a functional evidence for a role in obesity/adipogenesis or in which common variants are associated with BMI or a related trait and therefore warrant further consideration in follow-up analyses.

PrgmNr 3277 - Whole Genome Sequence Analysis of Dyspnea in Current and Former Smokers in COPDGene

View session detail

Author Block: **J. Chiles**[1], H. K. Tiwari[2], M-L. N. McDonald[1]; [1]Univ. of Alabama at Birmingham, Birmingham, AL, [2]Univ Alabama at Birmingham, Birmingham, AL
Disclosure Block: **J. Chiles:** Major Stockholder/Ownership Interest; Amgen.
INTRODUCTION: The development of dyspnea, both in patients with and without COPD, is known to have significant effects on quality of life for those who suffer from it. Various risk factors for the development of clinically-significant dyspnea have been identified and the development of dyspnea itself is a predictor of five-year mortality. Identification of genomic regions associated with dyspnea could provide mechanistic insight as to why only some patients develop dyspnea. **METHODS:** Dyspnea was dichotomized based on whether a participant reported a modified Medical Research Council (mMRC) dyspnea score greater or equal to 2. Trans-Omics for Precision Medicine (TOPMed) Freeze 8 whole-genome sequencing (WGS) data from 8390 current and former smokers from COPDGene were analyzed. Single nucleotide variants (SNVs) with minor allele frequency > 0.1% were analyzed in 5594 Non-Hispanic White (NHW) and 2796 African American (AA) participants, separately. SNVs were tested for association with dyspnea using SAIGE adjusting for age, sex, pack-years of smoking, and the presence or absence of COPD (defined as a Z-score of less than -1.64), along with 4 principal components of ancestry for the NHW and 6 principal components of ancestry for the AA. Analyses were performed using the NHLBIâ□□s BioData Catalyst cloud-based environment. Physical positions for SNVs are reported using the Human Genome Reference 38. **RESULTS:** No SNV was associated with dyspnea at a level reaching genome-wide significance (P-8) in either the NHW or AA. In the NHW, the most statistically significant finding was for the 21: 19705527 SNV (Beta(95% confidence interval) = 5.71(4.66 - 6.75), P = $5.2x10^{-8}$). The 21: 19705527 SNV was intergenic with the nearest gene (1.3Mb) being Neural Cell Adhesion Molecule 2 (NCAM2). Multiple other loci in this region on chromosome 21 had p-values in the range of 10^{-6} to 10^{-7}. Additionally, several other SNVs in the chromosome 21 region as well as multiple regions on chromosomes 1, 5, 8, 12, and 13 enriched with suggestive p-values. In the AA, the most statistically significant finding was for the 3: 172861520 SNV (Beta(95% confidence interval) = -0.45(-0.36 - -0.54), P = 1.8×10^{-7}). **CONCLUSION:** Though no SNV achieved genome-wide significance in our analyses, there are multiple suggestive association signals meriting further follow-up in a larger sample size available in other studies included in the TOPMed consortium. Also, we will perform additional analyses on specific subset of the population of interest and further characterization of the full cohort with rare variant testing.

PrgmNr 3278 - Association of *CXCR6* with COVID-19 severity: host genetic factors underlying transcriptomic regulation in *3p21.31* locus

View session detail

Author Block: **y. dai**[1], J. Wang[2], H-H. Jeong[1], W. Chen[3,4], P. Jia[5], Z. Zhao[6]; [1]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]MD Anderson Cancer Ctr. UTHlth.Graduate Sch. of BioMed. Sci., Houston, TX, [3]Immunobiology and Transplant Sci. Ctr., Dept. of Surgery, Houston, TX, [4]Dept. of Surgery, Weill Cornell Med., Cornell Univ., New York, NY, [5]7000 Fannin St., Houston, TX, [6]Univ Texas HSC Houston, Houston, TX

Disclosure Block: **Y. dai:** None.

Background: The coronavirus disease 2019 (COVID-19) is an infectious disease that mainly affects the host respiratory system with ~80% asymptomatic or mild cases and ~5% severe cases. Recent genome-wide association studies (GWAS) have identified several genetic loci associated with severe COVID-19 symptoms, such as *3p21.31* locus. There is a cluster of 6 genes (*SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6*, and *XCR1*) nearby the lead SNP rs35081325, which makes the â□□causalâ□□ gene and functional implication of this locus remain elusive. **Methods**: We implemented integrative approaches, including transcriptome-wide association studies (TWAS), colocalization and functional element prediction analyses, to interpret the genetic risks in lung, whole blood, and immune cells using two independent GWAS datasets from Host Genetics Initiative round 4 A2 (2,972 severe respiratory confirmed COVID-19 cases and 284,472 controls with unknown SARS-CoV-2 infection status) and Severe COVID-19 GWAS Group (1,980 COVID-19 confirmed cases with severe status and 2,205 control participants). To understand the context-specific molecular alteration, we further performed deep learning-based single cell transcriptomic analyses on a bronchoalveolar lavage fluid (BALF) dataset from moderate and severe COVID-19 patients. **Results**: In TWAS analysis, we discovered and replicated the genetically regulated expression of *CXCR6* and *CCR9* genes. These two genes have a protective effect on lung and a risk effect on whole blood, respectively. The colocalization analysis of GWAS and *cis*-expression quantitative trait loci highlighted the regulatory effect on *CXCR6* expression in lung and immune cells. The Hi-C cell line data from lung also showed significant interactions between lead SNPs and promoter regions of both *CXCR6* and *CCR9*. In lung resident memory $CD8^+$ T (T_{RM}) cells, we found a 2.24-fold decrease of cell proportion and lower expression of *CXCR6* (fold change = 0.56, two-sided Wilcoxon p = 1.8 Ã□ 10^{-18}) in severe patients than moderate patients. Pro-inflammatory transcriptional programs, apoptosis, and hypoxia pathways were highlighted in T_{RM} cells transition from moderate to severe groups. **Conclusion**: *CXCR6* tends to have a lower expression in severe patients than in moderate patients, which aligns with the protective effect of *CXCR6* from TWAS analysis. We illustrated one potential mechanism of host genetic variants or other unknown risks that might impact the severity of COVID-19 through altering the expression of *CXCR6* and lung T_{RM} cell proportion and stability, therefore, impairing the first-line defense in lung. Our results shed light on potential therapeutic targets for severe COVID-19.

PrgmNr 3279 - Asthma-associated genetic variants induce *IL33* differential expression through a novel regulatory region

View session detail

Author Block: **I. Aneas-Swanson**[1], D. C. Decker[2], D. R. Sobreira[1], N. Sakabe[1], K. M. Blaine[2], K. M. Magnaye[1], S. Clay[1], C. Ober[1], A. I. Sperling[2], M. A. Nobrega[1]; [1]Univ. of Chicago, Human Genetics, Chicago, IL, [2]Univ. of Chicago, Med., Section of Pulmonary and Critical Care Med., Chicago, IL
Disclosure Block: **I. Aneas-Swanson:** None.
Genome-wide association studies (GWAS) have implicated the *IL33* locus in asthma, but the underlying mechanisms remain unclear. We used the LD structure at this locus across populations of different ethnicities combined with a Bayesian fine-mapping tool to define a critical 20 kb genomic interval containing candidate causal SNPs for the asthma association. Epigenetic signatures further reduced this region to 5 kb, which we demonstrated to have enhancer blocking activity *in vivo* and *in vitro*. Chromatin conformation showed a crucial role in looping formation between *IL33* promoters and the regulatory elements within the critical interval. Humanized mouse BAC transgenics further confirmed the necessity of the 5 kb region for proper IL33 regulation. To identify nuclear proteins that can differentially bind to the risk and non-risk alleles and potentially alter *IL33* expression, we performed an electrophoretic mobility shift assay (EMSA) followed by mass spectrometry and selected OCT-1 (POU2F1) as a candidate differentially bound to the risk allele of rs1888909. RNA-seq data obtained from asthmatic and non- asthmatic subjects showed that carriers of one or two copies of the rs1888909 (T) risk allele had significantly higher *IL33* transcript levels in airway epithelial cells and IL-33 protein in the plasma compared to non- carriers of this allele. Our data demonstrate that asthma-associated variants at the *IL33* locus mediate allele-specific regulatory activity and *IL33* expression, through differential binding of OCT-1 to the asthma-risk allele. Altogether, these data provide a novel mechanism through which a regulatory SNP contributes to genetic risk of asthma.

PrgmNr 3280 - Complex Architecture of COVID-19 Host Genetics from Whole Genome Sequencing of a Multi-Ethnic Hospitalized Cohort

View session detail

Author Block: **D. M. Jordan**[1], L. G. Sloofman[1], R. C. Thompson[1], D. M. Del Valle[1], H. Shah[1], N. W. Simons[1], S. Gnjatic[1], M. Merad[1], E. Cheng[1], B. Losic[1], R. Sebra[2], N. D. Beckmann[1], A. W. Charney[1], E. Schadt[1], The Mount Sinai COVID-19 Biobank; [1]Icahn Sch. of Med. at Mt Sinai, New York, NY, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY

Disclosure Block: **D.M. Jordan:** None.

Several large studies have been performed on COVID-19 host genetics, and multiple relevant loci have been found. However, very few of these studies have examined features other than binary indicators of infection and severe disease. Additionally, these studies have been conducted primarily in subjects of European ancestry. We analyzed 12 traits measuring different features of COVID-19 disease progression and severity in a multi-ethnic cohort of 486 patients hospitalized for severe COVID-19 in the Mount Sinai Health System in New York, NY. These traits consisted of 4 quantitative measures of disease severity within the hospitalized cohort, 4 indicators of improvement or decline after hospitalization, 2 measures of intensity and effectiveness of immune response, and 2 measures of timing of disease progression. We collected whole genome sequences and performed multi-ethnic GWAS on common variants using the GENESIS package, which has been shown to be robust to heterogeneous ancestry and admixture. Despite the small sample size, we detected a locus at 9q33.2 associated with decline and end organ damage; three loci at 6p12.1, 9q33.1, and 21q21.2 associated with clearance of virus after hospitalization; a locus at 13q12.3 associated with improvement after hospitalization; and a locus at 10p15.1 associated with time between onset of symptoms and hospitalization. Many of the specific variants identified are not present on any standard genotyping array and could only be detected by whole genome sequencing. We also performed a SMMAT association test aggregating predicted loss of function variants by gene, also using GENESIS, and found two genes nominally associated with time between hospitalization and discharge or death: *MNDA* (FDR = 0.085), previously identified as a component of the interferon response; and *NPNT* (FDR = 0.059), previously implicated in lung disease including COPD and silicosis. Except for 9q33.1-2, which may be near the well-known *ABO* locus, none of these loci has been previously reported to be involved in COVID-19 host genetics. Additionally, again except for 9q33.1-2, each locus is unique to only one of the measured traits. Taken together, these results suggest that the host biology of COVID-19 infection is complex and multifactorial and is controlled by different factors at different stages, and binary indicators of infection and severity capture only a small fraction of host biology. These results also illustrate the enormous improvement in power that can be achieved using multi-ethnic cohorts, whole-genome sequencing rather than genotyping arrays, and granular, biologically meaningful phenotypes.

PrgmNr 3281 - Fine-mapping studies distinguish genetic risks for childhood- and adult-onset asthma in the HLA Region

View session detail

Author Block: **S. Clay**[1], N. Schoettler[1], A. Goldstein[1], P. Carbonetto[1], M. Dapas[1], M. C. Altman[2], M. G. Rosasco[2], J. E. Gern[3], D. J. Jackson[3], H. Im[1], M. Stephens[1], D. L. Nicolae[1], C. Ober[1]; [1]Univ. of Chicago, Chicago, IL, [2]Benaroya Res. Inst., Seattle, WA, [3]Univ. of Wisconsin, Madison, WI
Disclosure Block: **S. Clay:** None.
Genome-wide association studies (GWAS) of asthma have described robust associations with variation across the human leukocyte antigen (HLA) complex, including significant associations with both childhood-onset asthma (COA) and adult-onset asthma (AOA). This includes independent associations centered on the HLA class I (*HLA-C/B*) and class II (*HLA-DR/DQ*) regions for both COA and AOA in individuals from the UK Biobank (UKB). However, the specific HLA region variants and genes contributing to risk are unknown. In this study, we examined SNPs, HLA alleles, and HLA amino acid polymorphisms in British white individuals from the UKB. We used SuSiE to fine map both the COA and AOA risk loci and expression quantitative trait loci (eQTLs) in three asthma-relevant cell types, and examined putatively causal protein coding variation. In the class I region, our studies revealed a credible set (CS) containing a single, shared variant (rs2428494) for both COA (Posterior Inclusion Probability (PIP)=0.97) and AOA (PIP=0.999), and a COA-specific CS with two highly correlated variants: rs28481932 (PIP=0.43) and HLA-C Ala11Ser (PIP=0.57). In the class II region, however, there were no shared variants. Of the two COA CSs, one included one SNP (rs28407950, PIP=0.997) and one contained 5 non-coding SNPs spanning 152kb (median r^2=0.99). Two CSs were also identified for AOA. CS1 contained 60 variants (median r^2=0.99) and CS2 contained 33 variants (median r^2=0.96), with a mix of SNPs and HLA amino acid polymorphisms in both. eQTL fine mapping with SuSiE revealed that the same four putatively causal SNPs for AOA in CS1 (rs9272346, rs9274660, rs1063355, rs3828789) were also putatively causal for expression of the nonclassical class II genes, *HLA-DQA2* (P-9) and *HLA-DQB2* (P-10) in LCLs, PBMCs, and nasal epithelial cells and overlapped with strong enhancer annotations (H3K27ac) in LCLs from ENCODE. No variants in the class I or COA class II CSs were eQTLs. However, protein coding variation in HLA-C (Ala11Ser) for COA and the *HLA-DQA1**03 alleles in CS2 for AOA are plausible causal candidates. Overall, we highlight roles for both expression and protein coding variation in the HLA genes in asthma risk with both shared and distinct risk factors for COA and AOA, and provide a framework for how to conduct fine-mapping studies of variants in the HLA region. This research was conducted using the UK Biobank Resource under application number 44300.

PrgmNr 3282 - Infinity Biologix - A Biorepository Answers the Call and Becomes a Leading COVID-19 Assay Development and Testing Center

View session detail

Author Block: **M. H. Sheldon**[1], C. J. Bixby[1], C. P. Goswami[1], J. A. McDevitt[1], J. LaPorta[1], E. Kwon[1], P. Shah[1], J. Moore[2], C. Hevi[1], A. K. Bogdanowicz[1], R. H. Siegel[1], K. M. Nicolas[1], D. J. Murray[1], J. H. Schultz[1], J. W. Ruggieri[1], D. Garbolino[1], S. Nahas[2], R. Hager[1], R. Grimwood[1], A. I. Brooks[2]; [1]Infinity Biologix, Piscataway, NJ, [2]Infinity BiologiX, Piscataway, NJ
Disclosure Block: **M.H. Sheldon:** None.
RUCDR Infinite Biologics was established in 1998 and grew to become a world leader in providing the scientific community with comprehensive solutions in sample collection, processing, quality control, technical consultation, complex assays, storage and biosample distribution. On August 17, 2020, RUCDR Infinite Biologics became Infinity BiologiX (IBX, https://ibx.bio/). As IBX, we maintain the same core mission, to provide customers in the commercial and academic sectors the highest quality service and support in advancing their research. As a research organization and biorepository with a proven track record of innovation in developing and offering diagnostic assays, IBX was approached by local, state and federal entities to support the global community in expanding COVID testing by offering rapid testing to enable clinicians and patients to make critical decisions. In April 2020, IBX received the first FDA Emergency Use Authorization (EUA) enabling the use of saliva tests to detect the presence of the SARS-CoV-2 coronavirus, paving the way to the later approval of an EUA for its use in an at-home saliva sample collection kit. The IBX response to the pandemic was swift and decisive. Within a matter of months, significant investments were made in infrastructure and labor as well as collaborations established with two companies, Spectrum Solutions and Accurate Diagnostics Labs, to oversee the manufacturing and distribution of kits. As a result of these efforts, IBX established a COVID-19 PCR diagnostic lab with the capability of testing as many as 100,000 samples per day. Using the IBX TaqPath SARS-CoV-2 Assay (EUA200090), IBX has tested 7.5 million samples as of this writing, with 4.1% of those detected as positive and 95.2% coming in Spectrum saliva collection kits as opposed to nasopharyngeal swabs. Additional data will be presented that explores the global implications of our testing experience, including rates of positivity over time and detection of variants of concern in the US and internationally.

PrgmNr 3283 - Integrative analysis of DNase-I footprinting and genome-wide association data uncovers transcription factors involved in nine autoimmune diseases

View session detail

Author Block: **N. Hosseini Naghavi**[1,2], P. Shooshtari[3,1,2,4]; [1]Dept. of Computer Sci., The Univ. of Western Ontario, London, ON, Canada, [2]Childrenâ□□s Hlth.Res. Inst., Lawson, London, ON, Canada, [3]Dept. of Pathology and Lab Med., The Univ. of Western Ontario, London, ON, Canada, [4]Ontario Inst. for Cancer Res., Toronto, ON, Canada

Disclosure Block: **N. Hosseini Naghavi:** None.

Autoimmune and Inflammatory diseases (AIDs) are a group of > 80 complex diseases caused by loss of tolerance of the immune system for self-antigens. The biological mechanisms of AIDs are largely unknown, preventing the development of effective treatment options. Integrative analysis of genome-wide association studies and epigenetic data has shown that AID risk variants are enriched in epigenetic regions of immune cells, supporting their role in gene regulation. However, we still lack a systematic and unbiased identification of transcription factors (TF) involved in disease gene regulation.

We hypothesized that for some of the disease-relevant TFs, their binding to DNA is affected at multiple genomic sites rather than a single site, and these effects are cell-type specific. In this study, we developed a statistical approach to assess enrichment of TFs in being affected by disease risk variants at multiple sites. We used Immunochip data from nine AIDs and identified 99% credible interval (CI) SNPs for each trait. We then integrated CI SNPs and DNase-I footprinting data of 376 samples comprising 35 unique cell types, and used a probabilistic model (by Moyerbrailean et. al.) to identify the CI SNPs that are likely to change binding probability of certain TFs at specific cell types. Finally, for each TF (out of 1,372 TFs), we used Fisher's Exact test to assess whether CI SNPs are enriched in changing the binding probability of that TF at multiple sites (FDR Our analysis resulted in identification of significantly enriched TFs and their relevant cell types for each trait. The number of prioritized TFs in immune cell types varied between 1 and 14 for 7/9 AIDs. For the two other traits with smaller sample sizes, we did not find any significant TFs, likely due to a lack of statistical power. Our analysis identified some TFs previously known to be relevant to AIDs (e.g. Ahr:Arnt for rheumatoid arthritis and SPI-B for multiple sclerosis), and some other less studied new TFs. The enriched cell types also varied across the traits (e.g. CD8 and Mobilized CD4 T cells for rheumatoid arthritis, and CD56 and Mobilized CD4 T cells for multiple sclerosis). Our ChromHMM analysis proved that our predicted DNase-I footprinting sites are active enhancers or promoters in the relevant cell types. Additionally, our Great pathway analysis showed that the majority of the significant biological pathways are immune-related, an example of which is B cell adhesion pathway in multiple sclerosis. Our study provides a general framework for the identification of disease-relevant TFs and their relevant cell types, and facilitates discovering specific gene regulatory mechanisms of complex diseases.

PrgmNr 3284 - MHC class II alleles are associated with rare dermatologic autoimmune diseases in an analysis of over 200,000 participants of the UK Biobank

View session detail

Author Block: **D. T. Truong**[1], J. Molineros[1], Y. Chen[2], D. M. Waterworth[1], S. Li[3], M. Black[1]; [1]Janssen Res. & Dev., Spring House, PA, [2]Janssen Res. & Dev., San Diego, CA, [3]Janssen Res. & Dev., Spring house, PA
Disclosure Block: **D.T. Truong:** Salary/Employment; Janssen R&D.
Purpose: Dermatologic diseases are the fourth leading cause of disability worldwide and contribute to chronic pain, disfigurement, and poor mental health. Yet knowledge of disease pathogenesis remains unclear and treatment options are limited, especially for rare dermatologic autoimmune diseases (AuD) such as lichen planus (LP) and alopecia areata (AA). Variation in the major histocompatibility complex (MHC/HLA) accounts for nearly half of the explained genetic variation in AuD, highlighting its critical role in disease etiology. However, the role of HLA in rare dermatologic AuD is poorly understood, with genetic analyses to date limited to family-based or small-scale clinical studies. **Methods:** We curated 37 AuD phenotypes in the UK Biobank, a population-based cohort of ~500,000 individuals, using diagnostic codes (ICD9 and ICD10), primary care and self-reported data. We identified 3,610 and 886 cases of LP and AA, respectively. A common control set was defined as individuals with no self-reported or clinical diagnosis of any AuD (n=208,974). HLA imputation was previously conducted using HLA*IMP:02 in the UK Biobank, but recent comparison of different HLA imputation methods revealed suboptimal performance. We re-imputed HLA-A, -B, -C, -DPB1, -DQA1, -DQB1, and -DRB1 alleles from genotyping data to four-digit resolution using HLA Genotype Imputation with Attribute Bagging v4.1. Analyses were restricted to individuals of white British ancestry. Experimental models to evaluate the additive effect of HLA alleles on each AuD were fit using a Firth logistic regression to account for case-control imbalance. The reduced model did not include HLA. All models were adjusted for age, age^2, sex, age*sex, and the first 10 genetic principal components. We assessed HLA association using an omnibus test. **Results:** Novel associations were identified across several class II alleles for LP, including DQB1*5:01 and DQA1*1.01 (OR = 1.83, p = 5.98E-77 and OR = 1.76, p = 4.09E-61, respectively), as well as previously reported associations with DRB1*1:01 (OR = 1.83, p = 3.19E-53). For AA, a significant novel association was observed for class II allele DQB1*3:01 (OR = 1.36, p = 3.39E-7) and associations with DRB1*4:01 (OR = 1.49, p = 5.26E-6) were replicated. Residue and haplotype analyses and evaluation of utility for patient stratification are underway. **Conclusions:** MHC class II alleles are associated with increased risk of LP and AA, suggesting that genes in this region play a critical etiological role and provide a starting point to investigate promising biological mechanisms that could be targeted for future therapeutic development and improved treatment options.

PrgmNr 3285 - mRNA expression analysis reveals degenerative phenotype and immune cell infiltration in spleen, thymus and pancreas of *Clec16a* knockout (KO) mice

View session detail

Author Block: **M. A. Bakay**[1], B. P. Strenkowski[1], R. Pandey[1], H. Hakonarson[1,2]; [1]The Ctr. for Applied Genomics, The Childrenâ□□s Hosp. of Philadelphia, Philadelphia, PA, [2]Dept. of Pediatrics, the Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA
Disclosure Block: **M.A. Bakay:** None.
Multiple GWAS and functional studies prove *CLEC16A* as an important autoimmunity gene and suggest CLEC16A could be a critical regulator of aberrant autoimmune responses; however, the exact mechanism of its action is still largely unknown. To investigate the role of CLEC16A in autoimmunity we generated an inducible ubiquitous *Clec16a* KO mouse. The global *Clec16a* KO (*Clec16a*$^{\hat{I}\square UBC}$) exhibit a complex phenotype with marked immune dysfunction. We demonstrated loss of CLEC16A in immune cells leads to an abnormal mitophagy, cell death and immune dysfunction. Recently we reported dramatic pathological differences in spleen, thymus and pancreas of Clec16a$^{\hat{I}\square UBC}$ mice not observed in controls. We performed a longitudinal mRNA expression analysis using TaqMan RT-PCR expression assays. Considering the fast-evolving symptoms after knockout induction, we performed a time course expression study of the 16 immune markers in spleen, thymus and pancreas and observed a wide dysregulation of the expression in *Clec16a*$^{\hat{I}\square UBC}$ mice in comparison to controls. To exclude effect of tamoxifen (Tam) on the expression profile we have had two control groups (Oil and Tam) and observed no differences in them. The most significant findings were the following: CD163 expression was increased in all studied organs, suggesting macrophage infiltration. CD3, CD4 and CD8 markers were dysregulated in spleen and thymus. Adhesion proteins were dysregulated in spleen and thymus: Vcam1 up, Icam1 down. Bcl2, apoptosis regulator, was upregulated. FoxP3 and GzmB were upregulated in thymus only. While these finding require further in-depth investigation, they are in concordance with *CLEC16A* being a well-documented type 1 diabetes susceptibility gene and provide additional support for the direct involvement of CLEC16A in regulation pancreas, spleen and thymus functions. To determine predominant types of pancreas infiltrating immune cells and state of their activation we will perform high-resolution immunophenotyping. Identification of these cell subpopulations will inform on our choose of drugs aimed at reversing the abnormal processes in *Clec16a*$^{\hat{I}\square UBC}$ mice, that may be effective in treating and preventing symptoms of autoimmune disorders, such as T1D, in individuals with risk associated *CLEC16A* variants.

PrgmNr 3286 - Role of theFYVE and Coiled-Coil Domain Autophagy Adaptor 1 in severity of COVID-19 infection

View session detail

Author Block: **J. Shinn**[1,1], S. P. Smieszek[1,1], B. P. Przychodzen[2,1], M. H. Polymeropoulos[2,1]; [1]Vanda Pharmaceuticals, Inc., Washington, DC, [2]Vanda Pharmaceuticals, Washington, DC
Disclosure Block: **J. Shinn:** Salary/Employment; Vanda Pharmaceuticals, Inc.
Coronaviruses remodel the intracellular membranes to form specialized viral replication compartments, such as double-membrane vesicles where viral RNA genome replication takes place. Understanding the factors affecting host response is instrumental to the design of therapeutics to prevent or ameliorate the course of infection.
We obtained samples for whole genome and viral genome sequencing analysis from hospitalized patients with confirmed COVID-19 infection participating in VANDA COVID-19 studies, the strongest severity region thus far reported in association with severe COVID-19: is 3p21.31 with lead variant rs73064425 (rs73064425, OR=2.14, discovery p=4.77 Ã 10^{-30}) reported by Pairo-Castineira et al., 2020. We were able to replicate this association in our cohort of severe hospitalized patients with confirmed COVID-19 infection. We focused on the coding variants that are in LD with the tagging variant. Based on LD analysis, we report three coding mutations within one gene in the region of FYVE and Coiled-Coil Domain Autophagy Adaptor 1 (*FYCO1*). We specifically focus on the functional role of *FYCO1* nonsynonymous variants (rs33910087, rs13079478, rs13059238). Functionally, *FYCO1* encodes a protein involved in vesicle transport and autophagy. It has been suggested as a key mediator linking ER-derived double membrane vesicles, the primary replication site for coronaviruses, with the microtubule network. Interestingly, rs33910087 is an eQTL for CXCR6 (GTEX) and based on GeneATLAS, the genotyped variant is highly significant modifier of monocyte percentage ($p=3.8479e^{-46}$). We report the associations between the region and clinical characteristics in this set of hospitalized COVID-19 patients with severe course of infection.
To discover potential, pharmaceutical agents capable of affecting transcriptional expression levels of *FYCO1* implicated in SARS-CoV and SARS-CoV-2 pathophysiology, we have screened 466 compounds belonging to 14 different therapeutic classes. We focused only on compounds that significantly downregulate the expression of *FYCO1*. The most significant candidate was indomethacin - an anti-inflammatory drug that could potentially downregulate *FYCO1* already shown in a clinical trial to be efficacious in treatment of COVID-19. We hypothesize that via its direct effects on efficiency of viral egress, it may serve as a potent therapeutic decreasing the replication and infectivity of the virus. Further clinical studies will be necessary to examine the therapeutic utility of indomethacin and other compounds, downregulating *FYCO1* in COVID-19 infection and other strains of betacoronaviruses.

PrgmNr 3287 - The long-term effect of SARS-CoV-2 infection on host genetic regulation

View session detail

Author Block: **H-H. Chen**[1], D. M. Shaw[1], W. Zhu[1], L. E. Petty[1], M. Lee[2], J. B. McCormick[2], S. P. Fisher-Hoch[2], K. E. North[3], J. E. Below[1]; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Brownsville, TX, [3]Univ North Carolina, Chapel Hill, NC

Disclosure Block: **H. Chen:** None.

Since the COVID-19 pandemic, over 173 million individuals have been reported as confirmed infected cases globally. Although most of them successfully recover from the illness, some post-acute complications have been observed frequently, including fatigue, dyspnea, palpitations, and loss of smell and taste, and it may suggest COVID-19 is not only an acute infection but a chronic health condition. Therefore, the further and continued effect of SARS-CoV-2 infection is an urgent topic after the pandemic. In this study, we comprehensively investigated the transcriptome of post-infection subjects and aimed to study the long-term effect of COVID-19 on gene expression. We RNA-sequenced both prior and post COVID-19 whole blood samples of 81 antibody-confirmed cases from an established cohort, Cameron County Hispanic Cohort, and blood samples of two-time points from 88 age-, sex-, and interval-matched healthy controls. A mixed-effects model was applied to identify the genes with different changes in expression over time from prior to post-infection as compared to changes in expression over time in controls with no evidence of COVID-19 infection. Our preliminary analysis included complete two timepoints data from 27 cases and 48 controls, and five genes show a significantly different trend with p-value -4, including *LSM6* (p-value=$1.9x10^{-5}$), *GPR4* ($2.5x10^{-5}$), *RELN* ($2.9x10^{-5}$), *C8orf86* ($3.8x10^{-5}$), and *FUNDC2P2* ($9.5x10^{-5}$). *GPR4* encodes a G protein-coupled receptor that is highly expressed in the lung, heart, and kidney. Functionally, *GPR4* has been reported to regulate the inflammatory response and leukocyte infiltration and demonstrates an up-regulated trend in our post-COVID-19 samples which may suggest an important role for host immune response post-infection. Also, we used the 145 genes with p-value -3 for the phenome-wide enrichment test, and the secondary malignancy of respiratory organs is the most significant trait associated with our identified genes (p-value-5). The full analysis will include data from 338 RNA sequencing samples from two-time points in 81 cases and 88 controls and identified genesâ phenome-wide effects would be explored in Vanderbiltâs electronic health record-linked biobank.

PrgmNr 3288 - Trans-ethnic association analysis of HIV viral load and Chromosome X markers in individuals living with HIV

View session detail

Author Block: **C. I. Vergara**[1], P. McLaren[2], J. Tuff[3], The International Collaboration for the Genomics of HIV (ICGH), P. Duggal[4], E. Scully[1]; [1]Johns Hopkins Univ, Baltimore, MD, [2]Univ. of Manitoba, Winnipeg, MD, Canada, [3]Natl. Microbiol. Lab., Winnipeg, MB, Canada, [4]Johns Hopkins Bloomberg Sch. of Publ. Hlth., Baltimore, MD

Disclosure Block: **C.I. Vergara:** None.

Host genetic factors contribute to outcomes of HIV-1 infection. The robust phenotype of HIV-1 set point viral load (spVL), defined as $\log_{10}$ HIV-1 copies/mL blood of individuals during chronic, untreated infection, is a strong correlate of rate of disease progression and transmission potential and has been used to identify genetic determinants relevant to disease. Prior genome wide association studies (GWAS) identified genomic regions in the HLA class I region on chromosome 6 and the chemokine (C-C motif) receptor gene cluster on chromosome 3 that associate with disease phenotype. There are significant HIV-1 phenotype differences between men and women; women have ~0.4 $\log_{10}$ lower spVL as compared to men, and are more frequently spontaneous viral controllers. The genetic mechanisms for the observed sex differences are unknown. There is a paucity of data on sex-specific autosomal genetic architecture and on sex chromosome polymorphisms in HIV. We investigated the association of HIV spVL with markers on chromosome X in 10 879 individuals of European (N=7159, 6341 males/ 818 females) and African ancestry (N =3720, 2484 males/1235 females) enrolled as participants in the ICGH. 340,489 high quality genotyped and imputed variants on chromosome X from 17 genotype groups were tested for their association with HIV spVL. Linear regression models, with sex as a covariate and males coded as 0/2 copies of the minor allele (assuming X inactivation in females) were built for each group, and the results were combined using an inverse-variance weighted meta-analysis. Similar analyses were performed only in males. In the complete group, we found suggestive associations with the SNP rs6629425 (C>G, P $value_{Meta}=1.3x10^{-5}$) located ~3kb upstream of the Membrane Bound Transcription Factor Peptidase, Site 2 (*MSTP2*) gene. Each copy of the G allele has a decreasing effect in spVL consistent across most ancestry groups (effect sizes= -0.17 to -0.03). Male specific analysis showed similar, stronger results (rs6629425 P $value_{Meta}=1.8x10^{-6}$) with decreasing effect of G allele in spVL. *MBTPS2* is a membrane-embedded zinc metalloprotease that linked to sterol transcriptional control and endoplasmic reticulum stress response. Other suggestive associations were observed in the gene Neuroligin 4 X-linked (*NLGN4X*, rs5916333, P $value_{Meta}=2.8x10^{-5}$) and in an intergenic region (rs5986678, P $value=1.0x10^{-4}$). This analysis of X chromosome variation highlights potential genes and regions with effect on the variance of HIV spVL. Our findings are the first step towards a wider analysis evaluating the sex-specific genetic architecture of HIV related phenotypes in individuals of different ancestries.

PrgmNr 3289 - Whole-genome sequencing association study reveals methylation-mediated genotype association with decreased lung function in urban children

View session detail

Author Block: **M. Dapas**[1], E. E. Thompson[2], W. Wentworth-Shields[1], S. Clay[1], P. LeBeau[3], R. Z. Krouse[3], R. A. Wood[4], R. Cohen[5], R. G. Robison[6], G. K. Khurana Hershey[1], C. Kercsmar[7], R. S. Gruchalla[8], M. A. Gill[8], A. H. Liu[9], H. Kim[10], M. Kattan[11], L. B. Bacharier[12], D. Rastogi[13], W. W. Busse[14], P. J. Gergen[15], C. M. Visness[3], G. T. O'Connor[16], D. L. Nicolae[17], J. E. Gern[14], D. J. Jackson[14], C. Ober[17]; [1]Univ. of Chicago, Chicago, IL, [2]The Univ. Chicago, Chicago, IL, [3]Rho, Durham, NC, [4]Johns Hopkins, Baltimore, MD, [5]Boston Med. Ctr., Boston, MA, [6]Lurie Children's, Chicago, IL, [7]Cincinnati Children's, Cincinnati, OH, [8]UT Southwestern, Dallas, TX, [9]Children's Hosp. Colorado, Aurora, CO, [10]Henry Ford Hlth.System, Detroit, MI, [11]Columbia Univ. Med. Ctr., New York, NY, [12]Vanderbilt Univ. Med. Ctr., Nashvile, TN, [13]Children's Natl. Hosp., Washington, DC, [14]Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, [15]Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, [16]Boston Univ., Boston, MA, [17]Univ Chicago, Chicago, IL

Disclosure Block: **M. Dapas:** None.

Trajectories of lung function in early life predict the subsequent development of both asthma and COPD. Numerous genetic associations have been discovered for measures of lung function in genome-wide association studies, but these studies have mostly been limited to adults of European descent. Identifying genetic variants associated with lung function in a multi-ethnic pediatric population may highlight ethnic-specific risk loci or yield more generalizable insights into the development of chronic lung disorders. In this study, we analyzed measures of lung function from the Asthma Phenotypes in the Inner City (APIC) and Urban Environment and Childhood Asthma (URECA) cohorts, which consist of children living in low-income neighborhoods in 10 U.S. cities. We performed whole-genome sequencing on 897 individuals (ages 5-17 yrs; 67% Black, 25% Hispanic; 66% with MD-diagnosed asthma) and tested for associations with covariate-adjusted, normalized residuals for percent predicted forced expiratory volume in 1 sec (FEV1) and the ratio of FEV1 to forced vital capacity (FEV1/FVC). Sequence read processing, quality control, and variant calling were performed using the Genome Analysis Toolkit suite. Genetic associations were evaluated using linear mixed models accounting for asthma, age, sex, body mass index, ancestry, and relatedness, including 14.1M variants with MAF â¥1%. No associations with FEV1/FVC were genome-wide significant, but FEV1 was significantly associated with SNPs in a large LD block across the *TDRD9* gene on chromosome 14 (lead SNP rs10220464, $p=1.19x10^{-9}$). Because altered DNA methylation in the *TDRD9* gene in newborns was previously associated with maternal smoking, we analyzed DNA methylation (EPIC array) in nasal epithelial cells from 246 URECA children (age 11) and tested for correlations with FEV1-associated variants. rs10220464 was an meQTL for cg03306306, a CpG site located 65kb from the lead SNP, also in *TDRD9* ($p=1.54x10^{-4}$, FDR-2). Together, these results highlight a novel methylation-mediated genetic association with lung function, potentially linking environmental exposure to genetic risk on pulmonary health in urban youth.

PrgmNr 3290 - APOL1 and biobanking in the West African terrain-challenges and successes

View session detail

Author Block: **E. Salia**[1], R. L. Davis[2], J. L. Browne[3], E. K. Srofenyoh[4], C. L. Simpson[1]; [1]Dept. of Genetics, Genomics and Informatics, Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [2]Ctr. for BioMed. Informatics, Dept. of Pediatrics, Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [3]Julius Global Hlth., Julius Ctr. for Hlth.Sci. and Primary Care,Univ. Med. Ctr. Utrecht, Utrecht Univ., Utrecht, Netherlands, [4]Greater Accra Regional Hosp., Accra, Ghana
Disclosure Block: **E. Salia:** None.
Preeclampsia is a pregnancy related condition that affects 4% of pregnancies in the United States of America (USA) and has a higher morbidity and mortality in many low- and middle-income countries (LMICs). As a multi-systemic syndrome, it affects the kidneys and many other organs. Variants in Apolipoprotein L1 (APOL1) gene, which confer resistance to Human Acquired Trypanosomiasis (HAT), have been implicated in kidney disease and more recently, preeclampsia. APOL1 variants are specific to people of African descent and are being extensively investigated as key predictors of preeclampsia. High prevalence of these variants in sub-Saharan African populations demands further investigation, and numerous studies are ongoing in Africa to explore the effects of APOL1 variants. SPOT-BIO is part of the Severe Preeclampsia Adverse Outcome Triage (SPOT) study, an ongoing international research collaboration with multiple sites in Ghana. We collect biological samples from mother/baby dyads to determine effects of APOL1 genotypes on risk of developing preeclampsia. With only limited research infrastructure existing in LMICs, we will describe challenges in setting up biobanks geared towards improving the quality of maternal and child health in Africa. We will delve into the feasibility of implementing biobanks in a West African country and explore what LMICs in the subregion need to focus on to build robust, sustainable biobanks to develop capacity for cutting-edge genetics research. We will also show that despite these challenges, we have already recruited over 450 women, making it the largest cohort of African women with early preeclampsia in the world.

PrgmNr 3291 - A Myasthenia Gravis genomewide association study implicates AGRN as a risk locus

View session detail

Author Block: **A. Topaloudi**[1], Z. Zagoriti[2], A. Flint[1], M. Martinez[1], Z. Yang[1], F. Tsetsos[3], Y-P. Christou[4], G. Lagoumintzis[2], E. Yannaki[5], E. Papanicolaou-Zamba[4], J. Tzartos[6], X. Tsekmekidou[7], K. Kotsa[7], E. Maltezos[8], N. Papanas[8], D. Papazoglou[8], P. Passadakis[8], A. Roumeliotis[7], S. Roumeliotis[7], M. Theodoridis[8], E. Thodis[8], S. Panagoutsos[8], J. Yovos[7], J. A. Stamatoyannopoulos[9], K. Poulas[2], K. Kleopa[4], S. Tzartos[2,10], M. Georgitsi[11], P. Paschou[1]; [1]Purdue Univ., West Lafayette, IN, [2]Univ. of Patras, Rio, Greece, [3]Democritus Univ. of Thrace, Alexandroupolis, Evros, Greece, [4]The Cyprus Inst. of Neurology and Genetics, Nicosia, Cyprus, [5]George Papanicolaou Hosp., Thessaloniki, Greece, [6]Tzartos NeuroDiagnostics, Athens, Greece, [7]Aristotle Univ. of Thessaloniki, Thessaloniki, Greece, [8]Democritus Univ. of Thrace, Alexandroupolis, Greece, [9]Univ of Washington, Seattle, WA, [10]Hellenic Pasteur Inst., Athens, Greece, [11]Aristotle Univ. of Thessaloniki, Thessaloniki, Greece
Disclosure Block: **A. Topaloudi:** None.

Myasthenia Gravis (MG) is a rare autoimmune disorder affecting the neuromuscular junction (NMJ). Here, we investigate the genetic architecture of MG via a genome-wide association study (GWAS) of the largest MG dataset analyzed to date. We performed GWAS meta-analysis integrating three different datasets (total of 1,401 cases and 3,508 controls). We carried out HLA fine-mapping, gene-based and tissue enrichment analyses and investigated genetic correlation to 13 other autoimmune disorders as well as pleiotropy across MG and correlated disorders. We observed the strongest MG association to *TNFRSF11A* (rs4369774, p=1.09Ã10 -13 ; OR=1.4). Gene-based analysis revealed *AGRN* as a novel MG susceptibility gene. HLA fine-mapping pointed to two independent MG loci: *HLA-DRB1* and *HLA-B*. MG onset-specific analysis reveals differences in the genetic architecture of Early-Onset (EOMG) versus Late-Onset MG (LOMG). Furthermore, we find MG to be genetically correlated with Type 1 Diabetes (T1D), Rheumatoid Arthritis (RA), late-onset Vitiligo and Autoimmune thyroid disease (ATD). Cross-disorder meta- analysis reveals multiple risk loci that appear pleiotropic across MG and correlated disorders. Our gene-based analysis identifies *AGRN* as a novel MG susceptibility gene, implicating for the first time a locus encoding a protein (agrin) that is directly relevant to NMJ activation. Mutations in *AGRN* have been found to underlie congenital myasthenic syndrome. Our results are also consistent with previous studies highlighting the role of the HLA and *TNFRSF11A* in MG etiology and different risk genes in EOMG versus LOMG. Finally, we uncover genetic correlation of MG to T1D, RA, ATD, and late-onset Vitiligo, pointing to shared underlying genetic mechanisms.

PrgmNr 3292 - Cognitive trajectories diverge by genetic risk in the Wisconsin Registry for Alzheimerâ□□s Prevention

View session detail

Author Block: **E. Vasiljevic**[1,2], R. L. Koscik[3,4], E. M. Jonaitis[3,4], T. J. Betthauser[5,4], S. C. Johnson[3,4,6], C. D. Engelman[7,3,4]; [1]Population Hlth.Sci., Sch. of Med. and Publ. Hlth., Univ. of Wisconsin - Madison, Madison, WI, [2]Ctr. for Demography of Hlth.and Aging, Univ. of Wisconsin - Madison, Madison, WI, [3]Wisconsin Alzheimer's Inst., Sch. of Med. and Publ. Hlth., Univ. of Wisconsin - Madison, Madison, WI, [4]Wisconsin Alzheimer's Disease Res. Ctr., Sch. of Med. and Publ. Hlth., Univ. of Wisconsin - Madison, Madison, WI, [5]Dept. of Med., Sch. of Med. and Publ. Hlth., Univ. of Wisconsin - Madison, Madison, WI, [6]Geriatric Res. Ed. and Clinical Ctr., Wm. S. Middleton Veterans Hosp., Madison, WI, [7]Sch. of Med. and Publ. Hlth., Univ. of Wisconsin - Madison, Madison, WI

Disclosure Block: **E. Vasiljevic:** None.

Background: *APOE* Îµ4 allele status predicts both age of late-onset Alzheimerâ□□s disease (LOAD) onset and rate of cognitive decline leading up to LOAD. However, most of those studies have been done in older samples and with LOAD diagnosis. There is less known about when *APOE* related differences in cognition emerge. Moreover, traditional modeling of *APOE* risk using Îµ4 allele carrier status does not account for non-linear risk between combinations of Îµ2, Îµ3, and Îµ4 alleles. Our objective was to identify if the rate of cognitive decline differs by the *APOE* score, a composite score of the six *APOE* allele combinations weighted by the log odds of their association with LOAD.

Methods: We analyzed cognitive trajectories (n=1074) across *APOE* scores of participants in the Wisconsin Registry for Alzheimerâ□□s Prevention (WRAP). WRAP is a longitudinal study of adults age 40-70 and non-demented at baseline who were followed for up to 16 years with assessments every 2 years. We compared the rates of decline across the *APOE* score for four neuropsychological composite scores using a mixed effects regression model for longitudinal change with age.

Results: We found a significant age by *APOE* score interaction in predicting cognitive decline such that the rate of cognitive decline for all four composite scores was greatest for people with the highest *APOE* score. The effect of age was non-linear on all outcomes. Age was modeled as a cubic polynomial for immediate learning, delayed recall, and PACC3 and a quadratic polynomial for executive function. For immediate learning and delayed recall, the age of detectable cognitive function difference by *APOE* score is between 60 and 65 years. The initial statistically significant (Îœ = 0.05) difference in predicted cognitive function between the lowest and highest risk *APOE* score is about 0.15 standard deviation (SD) for immediate learning and 0.24 SD for delayed recall. For executive function and PACC3, the age of detectable cognitive function difference by *APOE* score is between 55 and 60 years. The initial difference in predicted cognitive function between the lowest and highest risk *APOE* score is about 0.3 SD for executive function and 0.34 SD for PACC3.

Conclusions: These results indicate that the rate of cognitive decline is greatest for people with highest *APOE* risk. Mean cognitive function and decline divergence by *APOE* risk happens between the age of 55 and 65 in the WRAP sample.

PrgmNr 3293 - Concordance of genetic variation that affects neuroinflammatory biomarkers for Alzheimerâ□□s disease and that influences brain volumes

View session detail

Author Block: **Y. Jin**[1], P. Paschou[2]; [1]Purdue Univ., West Lafayette, IN, [2]915 W. State Street, West Lafayette, IN
Disclosure Block: **Y. Jin:** None.
The increased number of Alzheimer's disease (AD) is one of the most daunting and potentially costly implications of ever-longer life expectancies. Currently, treatments targeting amyloid-beta or tau-protein pathology all failed at their late stage of clinical trials. Neuroinflammation has been accepted widely as the third core pathology of AD over the years, which provides a new perspective to understand AD and develop new drugs or treatments. Neuroinflammation can also be a bridge to link amyloid-beta and tau protein pathologies of AD. However, the mechanisms of how neuroinflammatory biomarkers influence specific brain regions are still largely unknown. This work aims to apply imaging genetics technique to investigate the overlap of genetic influence on both neuroinflammatory biomarkers and volume of each brain region for AD.We performed genome-wide association study (GWAS) of neuroinflammatory biomarkers (87 AD, 277 mild cognitive impairment, 46 controls) and brain volume (153 AD, 256 mild cognitive impairment, 108 controls) changes at baseline and 12 months using the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. To examine genetic pleiotropy and concordance single nucelotide polymorphism (SNP) effect directions, GWAS summary statistics of three neuroinflammatory biomarkers, CD40 Ligand (CD40L), Epidermal Growth Factor (EGF) and Neutrophil Gelatinase-Associated Lipocal with significant SNP (p

PrgmNr 3294 - Creating sex-specific phenome risk classifiers to identify under-documented cases of developmental stuttering in electronic health records

View session detail

Author Block: **D. G. Pruett**[1], D. M. Shaw[1], H-H. Chen[2], L. E. Petty[3], H. Polikowsky[1], S. Kraft[4], R. M. Jones[1], J. E. Below[3]; [1]Vanderbilt Univ., Nashville, TN, [2]Nashville, TN, [3]Vanderbilt Univ Med Ctr., Nashville, TN, [4]Wayne State Univ., Detroit, MI
Disclosure Block: **D.G. Pruett:** None.
Introduction: Developmental stuttering is a speech disorder that typically begins around age 3 and is characterized by syllable/word repetitions and prolongations. Despite population prevalence of 1%, no population-based genetic studies of stuttering have been conducted and few, if any, studies utilizing electronic health records (EHRs). Given that communication disorders like stuttering are diagnosed by speech pathologists, not physicians, information regarding diagnosis of these disorders is not consistently noted in EHRs. Rather than relying on ICD 9/10 codes, medication lists, or other typical approaches to defining phenotypes in EHRs, keyword searches and text-mining algorithms followed by manual review have proved helpful in identifying stuttering. However, despite extensive manual review, many stuttering cases likely go undetected in EHRs. **Purpose:** The purpose of this project is to develop sex-specific phenome risk classifiers to identify high-likelihood cases of developmental stuttering in EHRs; this will enable use of large-scale DNA biobank resources for well-powered, sex-stratified GWAS of developmental stuttering, a highly-sex skewed disorder. **Method and Results:** A keyword text search for stuttering and related terms within the Synthetic Derivative, a de-identified EHR database at Vanderbilt University Medical Center (VUMC), identified ~18,000 individuals with at least one keyword mention in their records. From these records, 50 confirmed cases of males who stutter and 50 confirmed cases of females who stutter will be used to create a text-mining algorithm to identify high-likelihood cases and reduce manual review to ~1,500 files. Confirmed cases will be stratified by sex and matched to controls in a permutation-based, sex-stratified comorbidity analysis of phecodes. Significantly enriched phecodes will be used to create sex-specific phenome risk classifier prediction models to identify additional high-likelihood stuttering cases that lack clinical documentation and were not otherwise identifiable using the multi-step identification process described above. A previous study created and tested a non-sex specific phenome risk classifier with a positive predictive value of 83%, validating this approach for acquiring developmental stuttering cases (Pruett et al., 2021). Ultimately, the creation of the sex-specific phenome risk classifiers will increase the number of high-likelihood stuttering cases identified within BioVU, VUMCâ□□s biorepository, to conduct well-powered sex-specific genome-wide association studies to investigate the genetic etiology of this highly heritable and sex-skewed disorder.

PrgmNr 3295 - Genetic correlations between resilience and UK Biobank traits differ by biological sex

View session detail

Author Block: **J. Eissman**[1,2], L. Dumitrescu[1,2], E. R. Mahoney[1,2], S. Mukherjee[3], M. L. Lee[3], W. S. Bush[4], C. D. Engelman[5], Q. Lu[6], D. W. Fardo[7,8], E. H. Trittschuh[9,10], J. Mez[11], C. C. Kaczorowski[12], H. H. Saucedo[13], K. F. Widaman[14], R. F. Buckley[15,16,17], M. J. Properzi[17], E. C. Mormino[18], H-S. Yang[16,17], T. M. Harrison[19], T. Hedden[20], K. Nho[21,22], S. J. Andrews[20], D. Tommet[23], N. Hadad[12], R. E. Sanders[3], D. M. Ruderfer[2], K. A. Gifford[1], X. Zhong[6], N. H. Raghavan[24,25,26], B. N. Vardarajan[24,25,26], Alzheimerâ□□s Disease Neuroimaging Initiative, A4 Study Team, Alzheimerâ□□s Disease Genetics Consortium (ADGC), M. A. Pericak-Vance[27], L. A. Farrer[11,28], L-S. Wang[29], C. Cruchaga[30], G. D. Schellenberg[29], N. J. Cox[2], J. L. Haines[4], C. D. Keene[3], A. J. Saykin[22,31], E. B. Larson[3,32], R. A. Sperling[33], R. Mayeux[24,25,26], M. L. Cuccaro[27], D. A. Bennett[34], J. A. Schneider[34], P. K. Crane[3], A. L. Jefferson[1], T. J. Hohman[1,2]; [1]Vanderbilt Memory & Alzheimerâ□□s Ctr., Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Univ. of Washington, Seattle, WA, [4]Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, [5]Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, [6]Univ. of Wisconsin-Madison, Madison, WI, [7]Coll. of Publ. Hlth., Univ. of Kentucky, Lexington, KY, [8]Sanders-Brown Ctr. on Aging, Lexington, KY, [9]VA Puget Sound Hlth.Care System, Seattle, WA, [10]Univ. of Washington Sch. of Med., Seattle, WA, [11]Boston Univ. Sch. of Med., Boston, MA, [12]The Jackson Lab., Bar Harbor, ME, [13]UC Davis Alzheimerâ□□s Disease Res. Ctr., Univ. of California Davis Med. Ctr., Sacramento, CA, [14]Univ. of California, Riverside, Riverside, CA, [15]Melbourne Sch. of Psychological Sci., Univ. of Melbourne, Melbourne, VIC, Australia, [16]Ctr. for Alzheimerâ□□s Res. and Treatment, Brigham and Womenâ□□s Hosp./Harvard Med. Sch., Boston, MA, [17]Massachusetts Gen. Hosp./Harvard Med. Sch., Boston, MA, [18]Stanford Univ., Stanford, MA, [19]Helen Wills NeuroSci. Inst., Univ. of California Berkeley, Berkeley, CA, [20]Icahn Sch. of Med. at Mount Sinai, New York City, NY, [21]Ctr. for Computational Biology and Bioinformatics, Indiana Univ. Sch. of Med., Indianapolis, IN, [22]Indiana Alzheimer Disease Ctr., Indiana Univ. Sch. of Med., Indianapolis, IN, [23]Brown Univ. Sch. of Med., Providence, RI, [24]Columbia Univ., New York, NY, [25]The Taub Inst. for Res. on Alzheimerâ□□s Disease and The Aging Brain, Columbia Univ., New York, NY, [26]The Inst. for Genomic Med., Columbia Univ. Med. Ctr. and The New York Presbyterian Hosp., New York, NY, [27]John P. Hussman Inst. for Human Genomics, Univ. of Miami Sch. of Med., Miami, FL, [28]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [29]Penn Neurodegeneration Genomics Ctr., Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [30]Washington Univ. in St. Louis Sch. of Med., St. Louis, MO, [31]Indiana Univ. Sch. of Med., Indianapolis, IN, [32]Kaiser Permanente Washington Hlth.Res. Inst., Seattle, WA, [33]Massachusetts Gen. Hosp., Harvard Med. Sch., Boston, MA, [34]Rush Alzheimer's Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL

Disclosure Block: **J. Eissman:** None.

Approximately 30% of elderly adults are cognitively normal at time of death despite presence of Alzheimer's disease (AD) neuropathology at autopsy. Studying these â□□resilientâ□□ individuals provides a unique approach to studying AD. It is well-established that sex differences are present in AD, with growing evidence suggesting genetic factors contribute to such differences. To this end, we sought to elucidate the sex-specific genetic etiology of resilience to AD. We leveraged our groupâ□□s published latent variable model of resilience representing better-than-predicted cognitive performance given an individualâ□□s age, sex, and amyloid burden and performed sex-stratified meta-analyses across four cohorts of cognitive aging. Then we performed a series of genetic correlation analyses (GNOVA) between the sex-stratified resilience meta-analyses and sex-stratified UK Biobank (UKBB) GWAS, adjusting for multiple comparisons. We evaluated 3,592 traits tested in females and 3,448 traits tested in males. Regardless of sex, resilience was genetically correlated with

education and neuropsychiatric traits. Among males, we observed 37 statistically significant correlations, whereby 24 appeared to be male-specific with non-overlapping confidence intervals with females. For example, among males, we observed a positive correlation with pulse rate (Rho=0.095, CI=[0.052,0.138], P.Bonferroni=0.048), representing unhealthy heart-rate variability, but we did not observe this correlation in females (Rho:-0.034, CI:[-0.067,-0.003], P.Bonferroni=1.000). Among females, we observed statistically significant correlations with 31 traits, whereby 10 appeared to be female-specific with non-overlapping confidence intervals with males. For example, among females, we observed a negative correlation with taking sotalol, an arrhythmia medication, representing unhealthy heart-rate variability (Rho=-0.041, CI=[-0.058,-0.025], P.Bonferroni=0.003), but we did not observe this correlation in males (Rho=0.018, CI=[-0.002,0.039], P.Bonferroni=1.000). Although genetic predictors of healthy heart-rate variability were associated with higher resilience in females, such measures were associated with lower resilience in males. Together, these results highlight that biological pathways, such as cardiovascular pathways, may relate to resilience against the downstream consequences of amyloidosis in a sex-specific manner, and re-emphasize the importance of considering biological sex in genetic explorations of AD.

PrgmNr 3296 - Genetic variants in the *SHISA6* gene are associated with delayed cognitive impairment in two family datasets

View session detail

Author Block: **W. K. Scott**[1], J. Ramos[1], S. H. Slifer[1], L. J. Caywood[1], M. B. Prough[1], J. E. Clouse[1], D. A. Dorfsman[1], S. D. Herington[1], M. D. Fuzzell[2], S. L. Fuzzell[2], S. D. Miller[2], M. D. Osterman[2], L. R. Main[2], O. Laframboise[2], K. Miskimen[2], A. H. Lynn[2], P. L. Whitehead[1], L. D. Adams[1], R. A. Laux[2], Y. E. Song[2], A. F. Zaman[1], T. M. Foroud[3], R. P. Mayeux[4], P. K. Ogrocki[5,2], A. J. Lerner[5,2], P. H. St. George-Hyslop[6], J. M. Vance[1], M. L. Cuccaro[1], J. L. Haines[2], M. A. Pericak-Vance[1]; [1]Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [3]Indiana Univ. Sch. of Med., Indianapolis, IN, [4]Columbia Univ. Med. Ctr., New York, NY, [5]Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH, [6]Univ. of Toronto, Toronto, ON, Canada

Disclosure Block: **W.K. Scott:** None.

Studies of cognitive impairment (CI) in the Amish identified sibships containing multiple CI and cognitively normal (CN; unaffected after age 75) siblings. We hypothesize that CN people carry alleles delaying age at onset (AAO) of CI, preserving cognition in older age despite increased genetic risk. We conducted a genome-wide study (GWAS) to identify loci associated with AAO of CI.1,522 Amish aged 43-99 (mean age 73.1, 42% men) screened at least once for CI using the modified mini-mental status exam (3MS) were genotyped using Illumina chipsets. Genotypes were imputed for 7,815,951 single nucleotide variants (SNV) with minor allele frequency (MAF)>1%. Disease-free time, defined as 1) age at first 3MS result indicating impairment (AAO; 3MS 86, 1,160 CN individuals) was the outcome. Cox mixed-effects models examined association between age and each SNV, adjusting for sex, population structure, and kinship. Analysis stratified by *APOE* genotype evaluated effect modification by APOE-4. To replicate genome-wide significant findings, SNVs 500 kb on either side of the peak SNV were examined for association with age in the NIA-LOAD family dataset (1,785 Alzheimer disease, 1,565 unaffected, mean age 73.5). Three SNVs were significantly associated (p-8) with AAO in the Amish, on chromosomes 6 (rs146729640; Hazard Ratio (HR)=6.38), 9 (rs534551495; HR=2.82), and 17 (rs14538074; HR=3.35). Each region found the common allele associated with later AAO. Replication analysis detected association at rs146729640 on chromosome 17, with nominal statistical significance (HR=1.49, p=0.02). The effect at rs146729640 was stronger in the APOE-4 negative subsets in both the Amish (HR=8.96) and NIA-LOAD (HR=1.82) samples. The replicated genome-wide significant association with AAO on chromosome 17 implicates the *SHISA6* gene, which is involved in post-synaptic transmission in the hippocampus. The analysis stratified by APOE-4 carrier status suggests that this effect is stronger in people without an APOE-4 allele. GWAS studies have not reported association in this region with AD risk or AOO, suggesting this might be a novel locus influencing onset of cognitive impairment, and highlights the power of using unique populations.

PrgmNr 3297 - Genomic and Transcriptomic-wide Analysis Identifies Novel Genetic Risk Loci and Prioritization of Therapies for Myasthenia Gravis

View session detail

Author Block: **R. Chia**[1], S. Saez-Atienzar[2], N. Murphy[1], A. ChiÃ²[3,4,5], C. Blauwendraat[6], the International Myasthenia Gravis Genomics Consortium, R. Roda[7], P. Tienari[8,9], H. Kaminski[10], R. Ricciardi[11], M. Guida[11], A. De Rosa[11], L. Petrucci[11], A. Evoli[12], C. Provenzano[13], D. Drachman[7], B. Traynor[14]; [1]NIH, Bethesda, MD, [2]NIH, Bethesda, MD, [3]Univ. of Turin, Turin, Italy, [4]Inst. of Cognitive Sci. and Technologies, Rome, Italy, [5]Azienda Ospedaliero Univ.ria CittalÌ della Salute e della Scienza, Turin, Italy, [6]NIA, Bethesda, MD, [7]Johns Hopkins Sch. of Med., Baltimore, MD, [8]Helsinki Univ. Hosp., Helsinki, Finland, [9]Univ. of Helsinki, Helsinki, Finland, [10]George Washington Univ., Washington DC, DC, [11]Univ. of Pisa, Pisa, Italy, [12]Catholic Univ., Rome, Italy, [13]UniversitÃ Cattolica del Sacro Cuore, Rome, Italy, [14]Natl. Inst Aging, Bethesda, MD

Disclosure Block: **R. Chia:** None.

Myasthenia gravis is a chronic autoimmune disorder caused by antibody-mediated destruction of receptors at the neuromuscular junction (NMJ) resulting in loss of communication between motor neuron and skeletal muscle. Genetics are likely to play a role as myasthenic patients have a higher risk of developing other autoimmune diseases. We performed a genome-wide association study (GWAS) and a transcriptome-wide association study (TWAS) to identify the genetic risks and candidate genes involved in disease etiology. The discovery cohort consists of 1,873 myasthenic patients with acetylcholine receptor antibodies and 36,370 healthy individuals, whereas 354 cases and 7,058 healthy individuals from the UK Biobank were included in the replication cohort. In addition to previous signals in *PTPN22*, *HLA-DQA1/HLA-B*, and *TNFRSF11A*, two novel signals in the acetylcholine receptor subunits genes, which are common antigenic target of the autoantibodies, were identified. From GWAS, the risk variant located in a promoter region of the cholinergic receptor nicotinic alpha 1 subunit (*CHRNA1*) gene on the forward strand increased risk by 1.57 ($p=3Â·07Ã10^{-8}$). TWAS identified two genes of which lower expression in skeletal muscle was predicted to increase disease risk: cholinergic receptor nicotinic beta 1 subunit (*CHRNB1,* $p=3Â·01x10^{-6}$, Z=-4Â·67*)* and epidermal growth factor receptor family of receptor tyrosine kinases (*ERBB2*, $p=5Â·63x10^{-7}$, Z=-5Â·00). Both *CHRNB1* and *ERBB2* are highly expressed at the NMJ and involved in the formation of functional acetylcholine complex. Thus, it is reasonable that lower expression of these proteins could impair neurotransmission. Onset-stratified analysis demonstrate that the genetic architecture is different between patients with early (CHRNA1, *CHRNB1* and *ERBB2* genes. In an unbiased approach, we identified several immunotherapies that may modify the disease progression based on their genetic profile. Our results demonstrate the power of genomics as a viable strategy for drug repurposing across diseases.

PrgmNr 3298 - Identification of novel genetic associations of epilepsy with functional validation in a zebrafish model

View session detail

Author Block: **T. Nagai**[1], E. Knapik[2], E. Gamazon[3,4,5]; [1]Human Genetics Graduate Program, Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Genetic Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]MRC Epidemiology Unit, Univ. of Cambridge, Cambridge, United Kingdom, [5]Clare Hall, Univ. of Cambridge, Cambridge, United Kingdom
Disclosure Block: **T. Nagai:** None.
Epilepsy is a neurological disease that affects millions of people around the world, and is defined by the occurrence of repeated, unprovoked seizures. While several different anti-epileptic drugs (AED) are approved to prevent or reduce the occurrence of these seizures, there is still a large proportion of epilepsy cases that are currently considered drug-resistant, limiting the treatment options available for those patients. Genomic analysis has been instrumental in identifying novel drug targets, which could potentially lead to an increased repertoire of AEDs for previously unknown genes and mechanisms of epilepsy and reduce the number of drug-resistant epilepsy cases. Genome-wide association studies (GWAS) of epilepsy were performed, but only few of the novel findings have been functionally validated. In our study we utilize the PrediXcan method, which combines the imputation of genetically regulated expression (GReX) and transcriptome-wide association studies (TWAS), and has shown to be effective at identifying novel disease-gene associations. However, rather than use single-tissue models of gene expression, we will apply joint-tissue imputation (JTI) models that leverage multi-tissue expression quantitative trait loci (eQTL) data from the Genotype-Tissue Expression study (GTEx) combined with epigenetic data from the ENCODE project and Roadmap Epigenomics. This allows the JTI models to incorporate tissue similarity into GReX imputation, producing better predictive performance. By using these JTI models with S-PrediXcan to analyze summary statistics of previous epilepsy GWAS data, our study has the potential to identify novel functional genes and mechanisms of epilepsy that have an increased likelihood of being causal. Once identified, these genes can then be validated in a high-throughput manner using a functional zebrafish seizure model. Zebrafish have proven to be a robust model capable of complex seizure phenotypes using pentylenetetrazol (PTZ), a seizure-inducing drug. Zebrafish can be easily genetically edited using CRISPR and their expression can be altered using morpholinos to generate models of newly identified genes. We will use these models to screen for genes that increase seizure susceptibility and perform molecular analysis to understand underlying mechanisms. By combining these computational and functional approaches, we can identify new causal genes associated with epilepsy.

PrgmNr 3299 - Linkage and association of cognitive resilience to chromosomes 4 and 18 in Midwestern Amish

View session detail

Author Block: **D. Dorfsman**[1,2], M. B. Prough[1], L. J. Caywood[1], J. E. Clouse[1], S. D. Herington[1], S. H. Slifer[1], L. D. Adams[1], R. A. Laux[3], Y. E. Song[3], A. Lynn[3], M. Fuzzell[3], S. L. Fuzzell[3], S. D. Miller[3], K. Miskimen[3], L. R. Main[3,4], M. D. Osterman[3,5], P. Ogrocki[6,7], A. J. Lerner[6,7], J. Ramos[1], J. M. Vance[1,2], M. L. Cuccaro[1,2], J. L. Haines[3,5], M. A. Pericak-Vance[1,2], W. K. Scott[1,2]; [1]John P. Hussman Inst. for Human Genomics, Univ. of Miami Miller Sch. of Med., Miami, FL, [2]The Dr. John T. Macdonald Fndn. Dept. of Human Genetics, Univ. of Miami Miller Sch. of Med., Miami, FL, [3]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [4]Dept. of Genetics and Genome Sci., Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [5]Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, Cleveland, OH, [6]Dept. of Neurology, Univ. Hosp. Cleveland Med. Ctr., Cleveland, OH, [7]Dept. of Neurology, Case Western Reserve Univ. Sch. of Med., Cleveland, OH

Disclosure Block: **D. Dorfsman:** None.

Inherited risk of Alzheimerâ□□s disease (AD) can be estimated by aggregating the effects of genome-wide risk loci into a genetic risk score (GRS). Individuals that are cognitively resilient (CR) are disease-free despite being at elevated risk of AD due to higher GRS and advanced age; studies of these individuals might identify genetic variants that protect against AD. The Midwestern Amish population is genetically and culturally homogenous, making it suitable for such studies. By integrating genetic risk of AD into a genetic linkage framework, we may better illuminate evidence of linkage and association in high-risk families exhibiting CR.

Subjects were classified as cognitively impaired (CI) or CR using the modified mini-mental status exam (3MS). Genotypes (Illumina platform) for 1,061 Midwestern Amish adults were imputed using the HRC reference panel, generating doses at ~ 7 million SNPs. Nonparametric linkage analysis was performed on sibships (n=120) with â□¥ 2 CR members using a panel of 2,167 SNPs (MAF â□¥ 0.3). Ordered-subsets analysis (OSA) can increase evidence of linkage in the presence of trait heterogeneity by arranging families in order of a trait-related covariate. Families were ranked from high-to-low genetic risk of AD using the mean GRS of CR sibship members. Effect sizes of 24 genome-wide late-onset AD loci were used to calculate GRS (Kunkle et al, 2019). Regions with significantly increased LOD* in the OSA-subset were examined by mixed model association testing in the overall sample. SNPs (MAF > 0.01) within 5 Mb of observed LOD* peaks were retained.

Mean family GRS=2.97 Â± 0.54. After OSA, significant LOD* increases were found on chromosome (chr) 4 (21 families with highest mean GRS, peak LOD*= 3.07, p=0.017, 67.55 Mb) and chr 18 (22 families with highest mean GRS, peak LOD*=2.65, p=0.021, 48.68 Mb). These regions do not contain known AD risk loci. 55,535 SNPs (MAF â□¥ 0.01) in the linked regions were analyzed. SimpleM was used to estimate the number of independent tests performed across both regions, resulting in a Bonferroni-corrected significance threshold of Î±=5.2 x 10-5. The strongest evidence of association was observed at rs28646178 (p=2.56 x 10-5, OR = 0.85, chr 18: 46.4 Mb), The strongest association on chr 4 was at rs72636894 (p=9.25 x 10-4, OR = 0.75).

These results suggest that novel genetic loci may promote CR, and that accounting for known AD genetic risk factors in linkage and association studies is an effective approach toward identifying novel protective alleles underlying CR.

PrgmNr 3300 - Measuring similarities and differences between EOAD and LOAD trough genome-wide association study, colocalization and heritability analysis

View session detail

Author Block: **E. Lucio da Fonseca**[1], M. Jean-Francois[1], C. Yang[2], J. Thulaseedhara Kurup[3], S. Hackleman Slifer[1], E. R. Martin[1], A. Naj[4], B. W. Kunkle[1], L. A. Farrer[5], J. L. Haines[6], R. P. Mayeux[7], G. D. Schellenberg[8], M. A. Pericak-Vance[1], M. Victoria Fernandez[2], C. Cruchaga[2], C. Reitz[3], G. W. Beecham[1], Alzheimer Disease Genetics Consortium; [1]Univ. of Miami, Miller Sch. of Med., John P. Hussman Inst. for Human Genomics, Miami, FL, [2]Washington Univ. Sch. of Med., St. Louis, MO, [3]Columbia Univ., The Taub Inst. for Res. on Alzheimerâ□□s Disease and the Aging Brain, The Gertrude H. Sergievsky Ctr. and Dept.s of Neurology and Epidemiology, New York, NY, [4]Dept. of Biostatistics and Epidemiology / Ctr. for Clinical Epidemiology and Biostatistics, Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [5]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [6]Dept. of Population & Quantitative Hlth.Sci., Cleveland Inst. for Computational Biology, Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [7]Dept.s of Neurology, Psychiatry, and Epidemiology, Gertrude H. Sergievsky Ctr., Taub Inst. for Res. on the Aging Brain, Coll. of Physicians and Surgeons, Columbia Univ., New York, NY, [8]Univ. of Pennsylvania Perelman Sch. of Med., Path & Lab Med, Stellar Chance, Philadelphia, PA

Disclosure Block: **E. Lucio da Fonseca:** None.

Introduction: Alzheimer disease (AD) is a degenerative brain disease, the most common cause of progressive dementia and the sixth leading cause of mortality in the United States. While early onset AD (EOAD, age at onset [AAO] 65), it is still unclear how much the two forms overlap, particularly with respect to their etiology.

Methods: Genetic data on 21,622 individuals from the Alzheimer Disease Genetics Consortium (ADGC) were used: (1,476 EOAD and 9,695 LOAD cases, and 10,451 controls). Single-variant association analyses were performed using logistic regression: full model (ancestry, sex, APOE-É□4 dosage, and SNP) and reduced model (ancestry and SNP). Evidence for colocalization was assessed to test if there is an association for both EOAD and LOAD, and if it is driven by the same causal variants. Locus zoom analysis was performed for significant index EOAD SNPs. The SNP heritability (h^2) and genetic correlation (rg) were estimated through LD score regression, considering the additional models: reduced + sex (ancestry, sex, and SNP) and reduced + APOE-É□4 (ancestry, APOE-É□4 dosage and SNP).

Results: Under full model, we identified three novel biologically plausible loci associated with EOAD: Chromosome 5 (*RP11-6N13.1*, rs15620839, $P=4.15x10^{-09}$); Chromosome 6 (*BACH2*, rs62408171, $P=3.67x10^{-08}$); and Chromosome 14 (*PRKCH*, rs186388694, $P=6.65x10^{-09}$). At each locus, the signal was supported by multiple associated SNPs. These genes are associated to chronic myeloid leukemia and immunodeficiency, cerebral hemorrhage, stroke, bipolar disorders, and hyperactivity. *BIN1* and *APOE* loci have posterior probabilities greater than 0.8, indicating that these two loci highly likely colocalized between EOAD and LOAD. The other three loci, *RP11-6N13.1*, *BACH2*, and *PRKCH*, failed colocalizing. Heritability analyses showed higher h^2 values for EOAD (h^2=0.30, 0.32, 0.34 and 0.33) than LOAD (h^2= 0.14, 0.18, 0.18, and 0.14), for full, reduced, reduced + sex and reduced + APOE-É□4 models respectively. Genetic correlation showed moderate genetic overlap between EOAD and LOAD only for full (rg=0.35, p=0.0283) and reduced + APOE-É□4 (rg=0.34, p=0.0261) models.

Conclusion: There is still a gap in our understanding of the similarities and differences between EOAD and LOAD forms. This study begins to address this by quantifying the amount of genetic overlap, identifying novel loci associated with EOAD, and confirming multiple other LOAD loci. We also identified three novel loci that are biologically highly plausible for EOAD, supported by a larger

heritability and suggesting a stronger polygenic effect in EOAD than LOAD.

PrgmNr 3301 - APOE gene associated with dementia related disorders and neuropsychiatric diseases in the Hispanic population

View session detail

Author Block: **C. Xu**[1,2], V. Padilla[1], K. Wang[3]; [1]UTRGV, Brownville, TX, [2]Univ. of Texas Rio Grande Valley, Brownville, TX, [3]West Virginia Univ., Virginia, WV
Disclosure Block: **C. Xu:** None.
Alzheimer's disease (AD), a common form of dementia, is commonly seen in the aging populations and is known to have a strong genetic component to its development. AD is one of the most common neurodegenerative disorders; it is a genetically and clinically heterogeneous disease. Specific demographic factors and genetic variants have been identified in non-Hispanic populations; however, a limited study was conducted in the Hispanic population. Therefore, we focused on investigating known gene, APOE, and demographic factors associated with the development of AD-related phenotypes and two psychiatric diseases (depression and anxiety) within the U.S. Hispanic population in our current study. A total of 1,382 subjects were collected from the Texas Alzheimer's Research and Care Consortium (TARCC, N=1,320) and the Initial Study of Longevity and Dementia from the Rio Grande Valley (ISLD-RGV, N=62). Questionnaires for demographics, medical history, and blood/saliva samples were collected. We genotyped the APOE gene. From the three APOE alleles, Îµ3 (90%), Îµ4 (21%), and Îµ2 (6%), the APOE Îµ4 allele showed association with AD (p=8.45E-7) in the Hispanic population. In addition, APOE Îµ4 allele was also associated with anxiety (p=4.84E-26), while the APOE Îµ3 allele showed an association with depression (p=0.002). We provide additional evidence in which APOE Îµ4 increased the risk for AD in Hispanics. For the first time, APOE Îµ4 and Îµ3 alleles show increased risks for anxiety and depression in Hispanics, respectively. Further research is warranted to confirm the current findings.

PrgmNr 3302 - Association between polygenic risk for substance use and music engagement across multiple cohorts

View session detail

Author Block: **P. Coleman**, R. L. Gordon, D. E. Gustavson; Vanderbilt Univ. Med. Ctr., Nashville, TN
Disclosure Block: **P. Coleman:** None.
Substance use (SU) is highly prevalent, particularly the use of nicotine and alcohol. Music engagement is often utilized as an independent coping strategy and clinical treatment for SU. Recent phenome-wide association findings from our group show decreased risk of tobacco use disorder in female musicians. Whether there is a genetic relationship between SU and music engagement is unknown, though recent findings from our group show genetic correlations between rhythmic ability and increased tobacco use, contrary to the negative phenotypic association. This study aims to (a) explore the phenotypic and genetic associations between music engagement and SU, and (b) create a model of the interaction between music engagement, genetic risk, and SU behaviors. Preliminary phenotypic analyses show that music engagement is generally correlated with reduced SU, though some music behaviors (e.g., going to many concerts) are associated with increased SU. Using summary statistics from the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN) and UK Biobank, we constructed polygenic risk scores (PRS) for alcohol and nicotine use in the research database BioVU. PRS for smoking initiation, cigarettes per day, age of smoking initiation, and alcohol frequency were compared with musicianship status, which had been defined in a prior study via algorithmic categorization through text-mining of electronic medical records (e.g., 'guitarist' or 'musician') and validated through manual chart review. Here we found those with higher genetic risk for smoking initiation (UK Biobank) and higher number of drinks per week (GSCAN) were more likely to be musicians. To further explore the relationship between music engagement and SU, we plan to replicate these findings in two additional cohorts: the Adolescent Brain and Cognitive Development (ABCD) cohort, and the Vanderbilt University Online Musicality (OM) cohort, both of which include more detailed music engagement questionnaires and genotyping. Analyzing this relationship in ABCD will be particularly interesting, as it contains many nicotine and alcohol-related measures, and SU behaviors usually begin in adolescence. The present study uses cross-trait PRS analysis to test a recent framework by our group, which posits that musicians may be at a higher genetic risk for SU, but music engagement may actually reduce SU behaviors. This research may be evidence for the use of music interventions to reduce SU behaviors for people at high genetic risk, and showcases a cross-trait method to examine the genetic relationship between cognitive traits.

PrgmNr 3303 - Dimensional traits in genetic research of autism show greater heritability and generalization than case-control traits

View session detail

Author Block: **T. Thomas**, T. Koomar, J. J. Michaelson; Univ. of Iowa, Iowa City, IA
Disclosure Block: **T. Thomas:** None.

The complexity of the autismâ□□s phenotypic spectra is well-known, yet most genetic research in this area uses a binary case/control status as the target trait. Numerous instruments exist to measure the core features of autism, including social and communication difficulties (measured by SCQ), restricted and repetitive behaviors (measured by RBSR), and motor coordination problems (measured by DCDQ). Here, we employed the SPARK autism cohort to illuminate the genetic etiology of 15 scales of autism symptomatology (N = 6,904 to 7,961 participants with autism) by calculating SNP heritability, performing genome-wide association studies (GWAS), identifying pleiotropy with polygenic risk scores (PRS), and lastly replicating the genetic signal in the typically-developing ABCD cohort (N = 6,659). Surprisingly, the PRS derived from previous autism GWAS had no association with symptomatology. Likewise, despite previous reports of autism and educational attainment being positively genetically correlated, higher educational attainment PRS was significantly associated with less autism symptomatology in all three core domains. Other significant PRS associations include bipolar disorder and openness PRS positively correlated with higher DCDQ, while ADHD, major depression, and neuroticism PRS positively correlated with both higher DCDQ and higher RBSR. In addition, modeling the PRS-by-sex interaction revealed the autism PRS (PGC) having slightly positive or no effect in males, while in females the autism PRS was negatively correlated with the RBSR. Significant PRS-by-sex interactions were also found for bipolar disorder, anorexia, major depression, and several personality PRS, underscoring the central role of sex in mediating genetic risk. Significant SNP heritabilities were observed across many of the subscales, with RBSR subscales having the highest heritabilities, ranging from 0.11 to 0.25, followed by the DCDQ (0.07-0.14), and the lastly the SCQ (0.01-0.1). Finally, we tested the generalization of the genetic associations to ABCD traits by calculating PRS from the 15 subscale GWAS, which were then tested for association with behavioral and emotional problems measured by the Child Behavior Checklist (CBCL). Notably, PRS from our subscale GWAS outperformed the PGCâ□□s autism PRS in predicting all 8 CBCL subscales as well as total problems, despite our GWAS having much lower sample sizes than the PGC autism study. Overall, these analyses suggest that more meaningful phenotyping is at least as important as the pursuit of ever-larger sample sizes in the ultimate success of genetic research in neuropsychiatric conditions like autism.

PrgmNr 3304 - Identify high risk individuals for Alcohol use disorder using polygenic risk scores

View session detail

Author Block: **D. Lai**, COGA Investigators; Indiana Univ Sch. of Med., Indianapolis, IN
Disclosure Block: **D. Lai:** None.
Alcohol use disorder (AUD) is one of the most common and preventable health problems. The estimated heritability of AUD is ~50% and identification of high risk individuals for AUD based on their genetic constitution will help us develop novel and targeted prevention and therapeutic strategies. Polygenic risk scores (PRS) are sum of risk alleles across the entire genome and have shown promises in identification of high risk individuals for diseases such as cardiovascular disease, cancer, etc. However, current PRS of AUD can only explain up to 2.5% of the variation and cannot be used to evaluate disease risk. In this study, we aimed to develop new PRS of AUD that can be used to identify high risk individuals. We first meta-analyzed two large-scale genome-wide association studies (GWAS) of AUD and related problems: AUD identified using the ICD codes from the Million Veteran Program and Alcohol Use Disorder Identification Test (AUDIT) problems scores (items 4-10) from the UK Biobank. Both GWAS used European ancestry samples only (N_{total}=323,608). Due to the different phenotypes and cohorts studied, study-specific findings were excluded and 2,757,681 variants were used. Using the meta-analysis results as the discovery dataset, we calculated PRS in three target datasets: the Collaborative Study on the Genetics of Alcoholism (COGA, N_{case}=3,453, N_{contol}=1,616), Study of Addiction: Genetics and Environment (SAGE, N_{case}=630, N_{contol}=337), The Australian twin-family study of alcohol use disorder (OZALC, N_{case}=1,650, N_{contol}=527). All three target datasets were European ancestry only to match the discovery dataset. PRS-CS was used to calculate the posterior effect sizes and European ancestry samples from the 1000 Genomes Project was used as the linkage disequilibrium reference panel. Then PRS were calculated using PLINK. For COGA and OZALC, GEE framework was used with a random effect to adjust the family relationships. For SAGE, logistic regression was used. Sex and Birth-cohorts were used as covariates for COGA and SAGE while sex and age were used for OZALC. Since COGA samples were genotyped on different arrays, an array indicator was also included as a covariate. Compare to the remaining samples, those samples with the 10% highest PRS have Odds Ratio of 2.22 (95%CI: 1.73-2.84, P-value

PrgmNr 3305 - Identifying potential environmental influences on clinical comorbidities of schizophrenia through integration of electronic health data and genetics

View session detail

Author Block: **T. Vessels**[1], N. Strayer[1], K. Choi[2], T. J. Morley[3], H. Lee[2], J. W. Smoller[4], Y. Xu[1], D. M. Ruderfer[3]; [1]Vanderbilt Univ., Nashville, TN, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Massachusetts Gen Hosp, Boston, MA
Disclosure Block: **T. Vessels:** None.
Background: Clinical comorbidities affect 68% of patients with psychiatric disorders and contribute to 60% of premature mortalities. Patients with schizophrenia (SCZ) have a reduced life expectancy of 10-20 years that is often a byproduct of associated comorbidities. We hypothesize that comparing clinical comorbidities from electronic health record (EHR) data to genetic associations will provide insights into which comorbidities are effects of the SCZ genetic etiology vs consequences of treatment, behavior or other environmental factors. **Methods:** We defined clinical comorbidities from a logistic regression of pairwise comparisons of diagnostic billing codes (PheCodes), adjusting for demographic and other covariates from two health care systems (Vanderbilt University Medical Center, VUMC; Mass General Brigham, MGB). Comorbidities were calculated on 250,000 randomly selected patients from each site individually and then averaged together. SCZ polygenic risk scores (PRS) in 64,190 genotyped patients of European ancestries were generated and compared in a logistic regression to each PheCode. After comparing the overall clinical and PRS associations, a TOST test of equivalence was used to further evaluate top SCZ comorbidities on whether they showed a well-powered absence of association with the SCZ PRS. **Results:** Clinical comorbidities were highly consistent across both institutions overall (r=0.79) and for SCZ specifically (r=0.85). We show significant replication of comorbidities previously identified in the literature including epilepsy, diabetes, and atrial fibrillation (9/15 with p-118) and cross institution (r =0.49, p =$8.4x10^{-81}$). After multiple test correction, there were 59 significant comorbid associations with SCZ. Interestingly, 11 of the 59 comorbidities had significant equivalence of SCZ PRS between cases and controls indicating an absence of contribution to phenotypic comorbidity from common variant SCZ genetics. **Discussion:** These 11 comorbid phenotypes may represent examples caused by other risk factors including rare variation, disease or treatment consequences or medical practice. Our results also identify several positive control examples with expected temporal patterns including movement disorders that are a known byproduct of antipsychotics and hepatitis as an outcome of behavior such as substance use. We further identify other comorbid phenotypes that could be modifiable environmental risk factors and point to strategies to improve health outcomes for SCZ patients.

PrgmNr 3306 - Investigating shared genetic liability between recurrent major depression and cardiometabolic traits in East Asian populations

View session detail

Author Block: **E. Lancaster**[1], T. B. Bigdeli[2], K. S. Kendler[3], B. T. Webb[3], R. E. Peterson[3]; [1]Virginia Commonwealth Univ., Richmond, VA, [2]SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [3]Virginia Commonwealth Univ, Richmond, VA
Disclosure Block: **E. Lancaster:** None.
Introduction: Major depression (MD) is a common yet potentially debilitating disorder that is often comorbid with other diseases, such as hypertension, coronary artery disease, and diabetes, further contributing to the overall impairment associated with MD. Given the high global burden and increasing rates of both cardiometabolic disorders and MD, investigating the origins of these disease relationships remains pertinent. While causal mechanisms linking these phenotypes are unknown, the moderate heritability reported for MD and several common cardiometabolic conditions suggests that shared genetic factors may be influencing these observed disease associations. This study utilized results from a genome-wide association study conducted in the CONVERGE (China, Oxford and Virginia Commonwealth University Experimental Research on Genetic Epidemiology) cohort alongside summary statistics from Biobank Japan to estimate the genetic correlation between recurrent MD and 14 cardiometabolic traits. Given the sex-specific differences in the prevalence and presentation of these traits, sex-stratified genetic correlations will also be examined. **Methods**: CONVERGE recruited 11,670 Han Chinese women with the primary goal of investigating etiological factors related to recurrent MD. Recurrent MD cases (n= 5,303) had experienced 2 or more episodes of MD defined by the Composite International Diagnostic Interview (CIDI). Healthy controls (n= 5,337) had no history of psychiatric illnesses. LD score regression (LDSC) was used to estimate genetic correlations between MD and 14 cardiometabolic traits selected from Biobank Japan, including both disease outcomes (e.g., type 2 diabetes) and related quantitative traits (e.g., hemoglobin A1c levels). To investigate whether sex-specific shared genetic influences may exist, additional genetic correlations were performed using the female-only subset of Biobank Japan for those phenotypes with sex-stratified summary statistics available (n= 7). **Results**: Significant negative genetic correlations were identified between MD and systolic blood pressure (rg= -0.22, p= 0.011), BMI (rg= -0.18 p= 0.001), and type 2 diabetes (rg= -0.15, p= 0.010). Analyses using female-only summary statistics showed a significant relationship between BMI (rg= -0.25, p= 3 x 10â»â´) and type 2 diabetes (rg= -0.18, p= 0.009). **Conclusions**: Results from this study suggest that MD may only partially share common genetic risk factors with some cardiometabolic traits. Further research is needed to determine whether observed correlations between these phenotypes result from common environmental factors or shared rare genetic variation.

PrgmNr 3307 - Pilot targeted gene analysis of resting heart rate variability in individuals with anxiety disorders and healthy controls

View session detail

Author Block: **J. Tomasi**, A. J. Lisoway, C. C. Zai, G. Zai, J. L. Kennedy, A. K. Tiwari; Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada
Disclosure Block: **J. Tomasi:** None.
Background: Heart rate variability (HRV) is a heritable measure of the variation in milliseconds between consecutive heart beats, and reflects the ability of the parasympathetic nervous system (PNS) to deactivate the sympathetic nervous system. Low HRV has been linked to poor cardiac health in addition to reduced psychological health. Anxiety disorder, which is related to deficient PNS activity, is associated with low resting HRV, and HRV is decreased during anxiety states. HRV may thus indicate the bodyâ□□s reactivity to anxiety and ability to recover. Therefore, genetic variants related to HRV may provide insight on the underlying biology of anxiety.
Methods: Our pilot sample consists of 47 individuals (15 patients with anxiety disorder and 32 controls, age 21-38, 33 female, predominantly European ancestry). Resting HRV was recorded via photoplethysmography for 5 minutes using a wearable Empatica wristband. Blood volume pulse data was processed via Kubios HRV software, and HRV was calculated using the root mean square of successive beat-to-beat interval differences. Global Screening Arrays were used to capture genome-wide SNP data. We performed a targeted analysis of SNPs in 17 genes involved in autonomic nervous system regulation (e.g. HPA-axis, serotonin, renin-angiotensin, norepinephrine, acetylcholine) using linear regression with an additive genetic model. Hypothesis-free GWAS was performed on HRV as well. Covariates included age, sex, diagnostic status, and the first two principal components of ancestry.
Results: The average HRV was 50.89msÂ±27.35ms (22.03ms-135.86ms). In our targeted gene analysis, rs4570625 from *TPH2* (Î²=16.59, p(unadj)=0.02), and rs878886 and rs110402 from *CRHR1* (Î²=26.30, 11.76; p(unadj)=0.004, 0.05) were nominally associated with HRV. In the hypothesis-free analysis, three SNPs passed Bonferroni correction: rs76239887, rs118065261 (*AKAP6*), and rs34434064 (Î²=56.12, 70.49, 77.83; p(unadj)= 4.7e-08, 8.5e-08, 1.1e-07; p(adj)=0.02, 0.04, 0.05).
Conclusion: This preliminary study identified several SNPs of interest for HRV through both hypothesis-free and hypothesis-driven approaches. In previous studies, *TPH2* rs4570625 and *CRHR1* rs878886 and rs110402 have been associated with anxiety-related phenotypes, and *AKAP6* has been associated with general cognitive ability as well as cardiac function. Our study is limited by the very small sample size, which will be increased as recruitment continues. Given the potential utility of HRV as a physiological marker of anxiety disorder, the SNPs of interest identified here, if replicated, may aid the identification of pathological anxiety risk and novel treatment targets.

PrgmNr 3308 - Targeted sequencing of genes most likely to harbor rare causal variants in schizophrenia

View session detail

Author Block: **D. Liu**, Psychiatric Genomics Consortium, Schizophrenia Exome Meta-analysis Consortium, A. Charney; Icahn Sch. of Med. at Mount Sinai, New York, NY
Disclosure Block: **D. Liu:** None.
The Schizophrenia Exome Sequencing Meta-Analysis (SCHEMA) Consortium conducted the largest-so-far whole exome sequencing (WES) study in 24,248 schizophrenia (SCZ) cases and 97,322 controls. Their results implicated an excess burden of ultra-rare coding variants in cases and identified genes conferring substantial risk at exome-wide significance. Based on these results, the Psychiatric Genomics Consortium (PGC) has conducted targeted sequencing of the top SCHEMA risk genes in a multi-ethnic cohort of 12,196 schizophrenia cases and 11,053 controls. The aims are (1) to replicate the findings from SCHEMA in a large independent cohort; (2) to meta-analyze with SCHEMA to increase the power for risk gene discovery.
The targeted panel comprised 161 genes with a high prior probability of harboring excess rare disruptive variants in SCZ. Targeted sequencing was performed on the Ion Torrent platform, and variants were called and filtered using an in-house pipeline optimized specifically for Ion Torrent data. Variant annotation was performed using The Ensembl Variant Effect Predictor, and missense variants were further prioritized using MPC pathogenicity score. The composition of the final callset was comparable to several other sequencing datasets including SCHEMA. SCHEMA demonstrated that rare protein-truncating variants (PTVs) are concentrated in constrained genes in SCZ cases compared to controls (OR=1.26), and we replicated this enrichment across the 80 constrained genes included in the sequencing panel, while carefully controlling for population stratification and technical confounding (P = 0.0016; OR = 1.29, 95% CI = 1.10 - 1.53). Conversely, damaging missense variants did not show a higher burden in cases compared to controls in this cohort (MPC > 3, OR = 1.00, P = 0.99; MPC 2~3, OR = 1.03, P = 0.66). When restricting to the 40 panel genes that ranked among the top 200 in SCHEMA, PTVs were more strongly enriched (P = 0.0004; OR = 1.44, 95% CI = 1.18 - 1.78), and damaging missense variants also showed extra burden in cases (MPC > 3, P = 0.43, OR = 1.43; MPC 2~3, P = 0.02, OR = 1.28). We also demonstrated a contribution of complete knockout (homozygous or compound heterozygous damaging variants) to SCZ risk (OR=2.26, P = 0.002). A meta-analysis with SCHEMA using Stoufferâ□□s method implicated two additional genes in which rare disruptive variants confer significant risk for SCZ: AKAP11 with P = 8.9 Ã□10-7, and SRRM2 with P =7.1 Ã□10-7. The integration of the genome-wide common SNP data is underway, which will enable a comprehensive assessment of the relationship between rare variant burden, copy number variation and polygenic risk from common SNPs.

PrgmNr 3309 - *MS4A* gene cluster variant modifies CSF sTREM2 associations with AD biomarkers of pathology

View session detail

Author Block: **R. Weiner**[1,2], L. Dumitrescu[2,3], K. Blennow[4], H. Zetterberg[5,6,7,8], K. A. Gifford[2], K. R. Pechman[2], A. L. Jefferson[2,3], T. J. Hohman[2,3], and for the Alzheimer's Disease Neuroimaging Initiative; [1]Dept. of Pharmacology, Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Memory and Alzheimer's Ctr., Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Vanderbilt Genetics Inst., Vanderbilt Med. Ctr., Nashville, TN, [4]Dept. of Psychiatry and Neurochemistry, Sahlgrenska Academy, Univ. of Gothenburg, MÃ¶lndal, Sweden, [5]Dept. of Psychiatry and Neurochemistry, Inst. of NeuroSci. and Physiology, Sahlgrenska Academy, Univ. of Gothenburg, MÃ¶lndal, Sweden, [6]Clinical Neurochemistry Lab., Sahlgrenska Univ. Hosp., MÃ¶lndal, Sweden, [7]Dept. of Neurodegenerative Disease, UCL Inst. of Neurology, London, United Kingdom, [8]UK Dementia Res. Inst. at UCL, London, United Kingdom
Disclosure Block: **R. Weiner:** None.
Background: Soluble Triggering Receptor Expressed on Myeloid Cells-2 (sTREM2) and its transmembrane parent (TREM2) are microglial signaling machinery, both of which have been implicated in the modification of Alzheimerâ□□s disease (AD) pathology. Higher CSF sTREM2 levels at baseline are thought to be associated with protection in AD, whereas membrane-spanning 4-domains subfamily A (*MS4A*) gene cluster variants modify both AD risk and levels of CSF sTREM2 protein. We previously found high baseline CSF sTREM2 relates to increases in CSF amyloid-Î² (AÎ²) species as well as blood-brain barrier (BBB) marker, CSF/plasma albumin ratio, leveraging Vanderbilt Memory and Aging Project (VMAP) participants. It is unknown whether *MS4A* variants interact with sTREM2 to modify levels of AD pathology. **Methods:** Baseline CSF sTREM2 measurement in 155 VMAP participants (meanÂ±standard deviation age=72Â±6.33, male=67%, mild cognitive impairment=47%), was assessed using linear regression models adjusted for age, sex, education, and cognitive diagnosis. Models examined whether sTREM2 levels interacted with *MS4A*-associated SNPs on the following biomarkers of AD pathology: $AÎ²_{x-42}$, $AÎ²_{1-42}$, $AÎ²_{x-40}$, total tau, phosphorylated tau, and the CSF/plasma albumin ratio. Alzheimerâ□□s Disease Neuroimaging Initiative (ADNI) data was used to replicate interaction results with AÎ² species measured by mass spectrometry (N=399). **Results**: CSF sTREM2 protein levels were significantly higher in minor allele carriers of rs1582763 than non-carriers (p=0.004). Previously defined associations of sTREM2 and CSF/plasma albumin ratio, $AÎ²_{x-40}$, as well as $AÎ²_{x-42}$, by our group were modified by the presence of rs1582763_A minor allele (Î²=7.087e-4, p.FDR=0.026; Î²=-0.439, p.FDR=0.033; Î²=-0.093, p.FDR=0.026, respectively) whereby higher sTREM2 was most strongly associated with higher AÎ² species among non-carriers. There were no significant interactions of rs1582763 on tau outcomes. Replication of AÎ² species results in ADNI approached but fell short of statistical significance ($AÎ²_{1-40}$ p=0.159, $AÎ²_{1-38}$ p=0.334 and $AÎ²_{1-42}$ p=0.689). **Conclusion:** AD-protective *MS4A*-associated rs1582763_A is associated with increased levels of CSF sTREM2 in VMAP, and the presence of this allele interacts with sTREM2 levels selectively on AD biomarkers. This includes shorter AÎ² peptides and a BBB marker, and excludes the well-established association between sTREM2 and tau. This suggests genetic regulation by *MS4A* of AD pathophysiology may be closer linked to amyloid rather than tau associations. Larger sample sizes are needed to clarify the modifying role of *MS4A* variants on sTREM2 associations with AD pathology.

PrgmNr 3310 - Analysis of genome-wide associations, exposome and lifestyle factors of skin-aging using medical proxies in the UK BioBank

View session detail

Author Block: **M. Liu**[1], A. Abbasi[1], J. Lee[1], B. Riley-Gillis[1], N. Smaoui[1], J. F. Waring[1], M. Reppell[1], X. Zheng[2], S. Manson Brown[3], T. Cheng[3], R. Mehta[3], M. Alexiades[4,5]; [1]AbbVie Inc., North Chicago, IL, [2]AbbVie Inc, Lake Bluff, IL, [3]Allergan Aesthetics, Irvine, CA, [4]Yale Univ. Sch. of Med., New Haven, CT, [5]Dermatology and Laser Surgery Ctr. of New York, New York, NY
Disclosure Block: **M. Liu:** Salary/Employment; AbbVie Inc..
Skin, the largest organ of the human body, plays a first-line role in protection against environmental and physical harm. Skin changes represent human health and aging, but a system to comprehensively analyze the etiopathogenesis of skin aging remains elusive and must assess many variables. Extrinsic damage due to ultraviolet radiation or pollution exposure causes skin health deterioration and visible skin aging, and heritability analysis from twin studies have shown that additive genetics may account for 60% of the variation in skin aging. Lifestyle factors such as smoking, poor sleep or diet have also been shown to influence skin aging. In this study, the UK Biobank (UKB), a large well-phenotyped population database, was utilized to assess the association of environmental, genetic, and lifestyles factors with skin-aging related traits. We integrated self-reported data and curated primary care data from the UKB to conduct a well-powered analysis of skin-aging related phenotypes. A combination of medical codes was utilized as proxy information for skin aging-related phenotypes - sunspots (N = 31,211), and broken capillaries (N = 1,111). To investigate genetic factors, a genome-wide association study (GWAS) was run in a European-ancestry subset of the imputed UKB genotypes for each of those traits and 12 significant loci were found to be associated with a proxy of sunspots. We replicated 2 loci containing genes, *MC1R* and *IRF4*, which were previously reported, for more precise skin-quality scores such as Beagley and Gibson six-point rating system. Co-dependent environmental or lifestyle factors that may affect skin aging were also assessed. Preliminary analysis demonstrated a strong positive correlation between time spent outdoors and poor sleep with sunspots. The research presented here provides a system to assess the environment, lifestyle and genetic factors of skin aging research in one sample population. Our findings highlight the value of using proxy phenotypes in biobank to define and elucidate the etiopathogenesis of skin quality and skin aging. Ongoing work will explore additional skin aging-related phenotypes and Mendelian randomization to explore the effect of exposure on the skin. Future work aims to replicate and validate these medical proxies in additional population cohorts.

PrgmNr 3311 - Associated phenotypic and genotypic characteristics in a cohort of patients with infantile hemangiomas: a case series

View session detail

Author Block: **M. Frigeni**, D. F. Rodriguez-Buritica; Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX
Disclosure Block: **M. Frigeni:** None.
Infantile hemangiomas are the most common benign tumors of infancy, being present in approximately 2% to 3% of neonates and in 10% of all infants by 12 months of age. The majority of infantile hemangiomas are present as an isolated finding; however, in a small subset of patients they can be associated with other anomalies, such as in case of PHACE and LUMBAR associations. To date, there is a lack of studies analyzing the spectrum of phenotypic features that can be found in patients with non-isolated infantile hemangiomas. Here, 39,017 subjects from the DECIPHER database were screened for the presence of an infantile hemangioma. A total of 85 subjects were identified, and their phenotypic features analyzed. In 75% of the subjects, a congenital anomaly was present in association with the infantile hemangioma. About 60% of the anomalies identified involved the musculoskeletal system, followed by the genitourinary (35%), nervous (28%), and cardiovascular (20%) systems. It was also noted that 87% of the subject had at least one comorbidity reported, with most of them being neuro-developmental problems. Lastly, facial dysmorphism was observed in 76% of the subjects. Only 5% of the subjects had no additional finding other than the infantile hemangioma. As a next step, genotypic analysis of the 85 subjects was performed and uncovered a high prevalence of copy number variations on Chromosomes 9 and 4. Interestingly, there are at least 11 genes with crucial dose effect for vasculogenesis located on Chromosome 9. This finding warrants further investigation of these genomic regions, and could shed new light on the mechanisms responsible for infantile hemangiomas, whose genetic bases remain unknown.

PrgmNr 3312 - Association of whole-person polygenic component scores with Alzheimer's disease pathology

View session detail

Author Block: **A. Kharaghani**[1,2], M. Milic[1], D. A. Bennet[3], P. L. Jager[4], J. A. Schneider[3], L. Sun[2,5], D. Felsky[1,2,6]; [1]Krembil Ctr. for Neuroinformatics, Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [2]Dalla Lana Sch. of Publ. Hlth., Univ. of Toronto, Toronto, ON, Canada, [3]Rush Alzheimerâ□□s Disease Ctr., Rush Univ. Med. Ctr., Chicago, IL, [4]Ctr. for Translational and Computational Neuroimmunology, Columbia Univ. Med. Ctr., New York, NY, [5]Dept. of Statistical Sci., Univ. of Toronto, Toronto, ON, Canada, [6]Dept. of Psychiatry, Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **A. Kharaghani:** None.
Common neurodegenerative disorders, such as Alzheimerâ□□s Disease (AD), are highly heterogeneous and genetically correlated with many other age-related traits. Growing online compendia of GWAS summary statistics now enable phenome-wide assessments of SNP-based polygenic scores (PRS). We harnessed such summary statistics to describe the â□□whole personâ□□ genetic risk landscape in an elderly population, and test if meta-PRS scores derived from over 2,000 traits can improve risk prediction of AD and AD-related phenotypes. The Religious Order Study and Memory Ageing Project (ROS/MAP) are two ongoing longitudinal studies of ageing, involving annual cognitive testing and post-mortem neuropathological assessment. We imputed genotypes for 2,052 unrelated ROS/MAP participants of European ancestry using the TopMed Imputation panel and calculated PRSs for up to 2,312 heritable ($h^2 > 0.05$) and sex non-specific phenotypes, selected from the Pan-UKBB consortium, at 15 different P-value thresholds using PRSice v2.3.3. PRSs were corrected for fine population structure and omitted if they were constructed from fewer than five SNPs. At each P-value threshold, principal component analysis was applied to PRSs for all traits, with the resulting latent components representing â□□meta-PRSâ□□. We observed that the top meta-PRSs explained up to 5.7% of the variability in all PRSs at lower p-value thresholds (e.g. $5x10^{-8}$) and loaded most strongly onto cardiovascular and metabolic phenotypes. At higher P-value thresholds, prescriptions for pain killers were the major loadings of top meta-PRSs. In direct phenotypic association analyses, meta-PRSs representing all-cause dementia, ocular health, and cardiovascular disease were most strongly associated with rates of cognitive decline and levels of amyloid and tau neuropathology. The clustering of â□□whole personâ□□ genetic risk in this population offers a map of heterogeneity in genetic risk profiles for AD-related traits and reveals methodological influences on broad-scope polygenic correlation in the elderly population.

PrgmNr 3313 - Candidate causal variants affecting tissue-specific regulatory mechanisms in the GWAS catalog

View session detail

Author Block: **D. Hemerich**[1], Z. Wang[1], R. Loos[2,3]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]The Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]The Novo Nordisk Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, Denmark
Disclosure Block: **D. Hemerich:** None.
Genome-wide association studies (GWAS) have revealed thousands of variants associated with a multitude of diseases and traits, the vast majority located in the non-coding region of the genome. These variants likely affect disease risk by acting on regulatory elements in tissues critical for the disease or trait of interest. While most regulatory elements, such as enhancers, are shared across tissues, some are tissue-specific and regulate genes that are essential for the function of that tissue. Identifying genetic variants acting in enhancers that are unique to tissues implicated in the disease can be a first step towards prioritizing candidate causal variants and identifying regulatory mechanisms. Here, we created a data resource of high confidence enhancers that are specific to one of 33 cell-types. We intersected these tissue-specific enhancers with independent variants from the GWAS catalog and their proxies at high linkage disequilibrium. We identified 20,405 independent loci associated with 3,149 diseases and traits that affect tissue-specific enhancers. We also retrieved the nearest gene to each associated variant as well as the gene predicted by the Epimap project, as candidate target genes that can be affected by the associated variant potentially disrupting the tissue-specific enhancers. In addition, we included the probability of causality of variants overlapping tissue-specific enhancers generated using PICS.Every tissue-specific set of enhancers showed enrichment for variants associated with diseases that implicate these tissues. For example, variants associated to pulse pressure are enriched in artery-specific enhancers (p=1.2E-10), those associated with intelligence are enriched in brain-specific enhancers (p=8.7E-06) and those associated with electrocardiogram morphology are enriched in heart-specific enhancers (p=2.2E-08). We observed well-known genes such as *PCSK9* implicated in liver-specific enhancers by variants associated to cholesterol levels, as well as *TTN* implicated in heart-specific enhancers by variants associated to atrial fibrillation. Taken together, we created a data resource that can help with prioritizing genetic variants acting in tissue-specific enhancers for functional studies. We plan to make this resource available as an online portal, to aid researchers in elucidating the biological mechanisms of diseases.

PrgmNr 3314 - Curated catalog of 309 functional GWAS non-coding variants across 125 diseases: a systematic review

View session detail

Author Block: **A. J. Alsheikh**[1], S. Wollenhaupt[2], E. A. King[1], J. Reeb[2], S. Ghosh[1], L. R. Stolzenburg[1], S. Tamim[1], J. Lazar[1]; [1]Genomics Res. Ctr., AbbVie, North Chicago, IL, [2]Information Res., AbbVie, Ludwigshafen, Germany
Disclosure Block: **A.J. Alsheikh:** Salary/Employment; AbbVie.
Genome-wide association studies (GWAS) have revolutionized our understanding of genetic susceptibility to complex human diseases. With more than 90% of GWAS associated variants being non-coding, the most challenging aspect remains the functional validation of these variants and deciphering the biological processes that underlie the genetic associations. The extent of experimental wet lab validation studies of GWAS non-coding variants conducted across all human disease traits remains largely unknown and no curated resource is published. To address this gap, a systematic literature review of experimental validation studies published in Medline from 2007 to 2020 was conducted in accordance with the 2020 PRISMA guidelines. We used an initial set of manually identified index studies to build a broad systematic search identifying 36,676 articles. Using natural language processing and ontology-based text-mining of abstracts and metadata, we built filters for 7 eligibility criteria 1) primary research articles, and mention of 2) GWAS, or association, 3) any human disease, 4) any human gene, 5) non-coding context, 6) functional, causal, or regulatory variant or specific rsID, and 7) wet-lab experimental validation techniques reducing the search results to 1,454 articles that met all the 7 criteria. These were manually reviewed for the same eligibility criteria. For 875 eligible articles, we manually curated the validated variant rsID, variant class, target gene symbol, and associated disease per article. This data was cross validated using the GWAS catalog resulting in a final set of 286 articles. These articles report 309 experimentally validated functional non-coding variants regulating 255 genes across 125 human disease traits. The variants cover a variety of classes and regulatory mechanisms and primarily act through gene promoters, enhancers, silencers, insulators, microRNAs, and long non-coding RNAs. Our analysis quantifies progress that has been made in the last decade and identifies the beginning of validation of distal regulatory element variants. Although extensive research is needed until we understand the biological mechanisms of all GWAS loci, we believe that this curated resource will provide tremendous value for a broad base of scientists who pursue validation studies by providing concrete prior exemplars of validations studies for wide variety of non-coding variant types in all disease areas. Additionally, for computational biologists interested in developing tools for functional non-coding variant prediction that has been previously hindered by lack of a truth set of experimentally validated variants.

PrgmNr 3315 - Determining the Genetic risk of Complex Disorders in Global populations

View session detail

Author Block: **P. JAIN**, M. Martinez, P. Drineas, P. Paschou; Purdue Univ., West Lafayette, IN
Disclosure Block: **P. Jain:** None.
Complex disorders are caused by a combination of genetic, environmental and lifestyle factors, and their prevalence can vary greatly across different populations. This difference could be due to genetic or environmental influences and understanding its basis could help us identify which groups of people might have a higher risk of developing certain disorders. Here, we undertake a systematic evaluation of genetic risk across worldwide populations for 20 different disorders and explore how it relates to disease prevalence. GWAS summary statistics of 20 different disorders (cardiovascular, neurological, metabolic, autoimmune and psychiatric disorders) were used to estimate Polygenic Risk Scores (PRS) in 9 European and 24 worldwide reference populations. We then estimate the correlations between average genetic risk for each of the 20 disorders and their prevalence around the world. We find significant correlations between prevalence and PRS for 13 of the studied disorders with Obesity having the highest correlation (R2=0.73, corrected P-value=0.001). A clear variation in genetic risk is observed based on ancestry and we identify populations that have a higher genetic liability for developing certain disorders both within European and global regions. Compared to other populations, Africans have a higher genetic risk for Obesity, East Asians have a higher genetic risk for Type 2 Diabetes and Coronary disease, and Europeans have a higher genetic risk for Alzheimerâ□□s and moderate risk for psychiatric and Autoimmune conditions. These results can help shed light on the genetic architecture of complex disease around the world and could inform early intervention strategies for populations at higher risk for certain disorders, promoting equity in medical treatment.

PrgmNr 3316 - DNA-methylation quantitative trait loci in Asian populations

View session detail

Author Block: **Y. Xia**[1,2], T-T. Chen[3], S-C. Lin[4], C-Y. Chen[5], Y-C. Feng[6], T. Ge[7], H. Huang[2], Y-F. Lin[8]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Ctr. for Neuropsychiatric Res., Miaoli, Taiwan, [4]Natl. Hlth.Res. Inst.s, Miaoli, Taiwan, [5]Biogen, Cambridge, MA, [6]MGH; Broad Inst., Cambridge, MA, [7]Broad Inst., Boston, MA, [8]Natl. Hlth.Res. Inst.s, Taiwan, Zhunan Town, Miaoli County, Taiwan

Disclosure Block: **Y. Xia:** None.

DNA methylation (DNAm) is a key regulator of transcriptional regulation and is increasingly recognized as a contributor to health and disease. Population-level variation in DNAm is under strong genetic influence, according to twin and family studies. Several studies in a variety of tissues, including the brain and whole blood, have found widespread associations between common DNA sequence variants and DNAm. These DNAm quantitative trait loci (mQTL) are enriched in regulatory chromatin domains and transcription-factor binding sites, providing critical insights into the connection between the noncoding genome and human health and disease phenotypes. However, most QTL studies to date were conducted in the European populations, leaving the noncoding genome of non-European individuals unexplored and exacerbating the healthcare disparity.To alleviate this critical knowledge gap, we performed the largest mQTL analysis to date in the Asian populations. DNAm was profiled in 2,039 subjects of Han Chinese ancestry from the Taiwan Biobank using the Illumina EPIC array. All subjects were also genotyped using the Taiwan Biobank array, with 8,033,242 genomic variants available after quality control and imputation. We tested the association between genetic variants and variations in 7,115,522 DNAm sites using principal components as covariates to control the population structure, PEER factors to control unknown variables. We removed batched effect and position effect by ComBat. We used sex, age, smoke experience, and cell proportion as covariates. Using a stringent significance threshold controlling multiple testing (p

PrgmNr 3317 - Fibromyalgia as a Nexus of Multiple Liability Distributions

View session detail

Author Block: **A. B. Faucon**[1], A. Moscati[2], C. Arnaiz-YÃ©pez[1], K. Sundquist[3], G. M. Belbin[2], J. H. Cho[2], R. Loos[4], K. S. Kendler[5], L. K. Davis[6]; [1]Vanderbilt, Nashville, TN, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Lund Univ., Lund, Sweden, [4]The Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Virginia Commonwealth Univ, Richmond, VA, [6]Vanderbilt Univ Med Ctr., Nashville, TN
Disclosure Block: **A.B. Faucon:** None.
Introduction: Epidemiological estimates indicate fibromyalgia affects up to 8% of the population[1], is underdiagnosed[2], and frequently lasts a lifetime. The heritability of chronic pain conditions is estimated at 50%[3]. Small sample sizes have precluded finding significant genetic loci to date, despite investigations using methods including linkage[4] and candidate gene studies[5,6], genome-wide association methods[7], and exome sequencing approaches[8]. Thus, to more fully elucidate the genetic architecture of fibromyalgia, we examine sex differences and provide estimates of prevalence, comorbidities across the phenome, polygenic overlap between fibromyalgia and other traits, and perform a genome-wide association study (GWAS) of fibromyalgia from the largest sample to date.
Methods: Fibromyalgia cases were identified by the ICD10 code M79.7. Controls were identified by presence of any ICD10 code, together with the absence of ICD9 code 729.1 and ICD10 code M79.7. Ancestry classification was based on principal component clustering. We calculated polygenic risk scores (PRSs) for eight comorbid phenotypes with PRSice2[9] in BioVU (Vanderbilt biobank) and BioMe (Mount Sinai biobank). We fitted multivariable logistic regression models to test the association between fibromyalgia diagnosis and each component PRS while controlling for median age and genetic PCs 1-10. We perform GWAS using Saige v0.42.1[10]with PCs1-10, age, BMI, and sex, as covariates and meta-analyze using METAL.
Results: Swedish register data indicate enrichment among fibromyalgia cases of comorbidities including major depression, anxiety, back pain, inflammatory bowel disease, asthma, rheumatoid arthritis, osteoarthritis, and decreased subjective well-being. In our analyses, fibromyalgia case status was significantly associated with PRSs for back pain, inflammatory bowel disease, major depression, neuroticism, rheumatoid arthritis, and subjective well-being. The PRS for rheumatoid arthritis demonstrated the strongest positive association with fibromyalgia (OR per standard deviation of risk: 1.9-2.2) while subjective well-being yielded the strongest negative association (OR: 0.84-0.91). The first genome-wide significant locus is identified by two separate SNPs and will also be reported.
Discussion: In this work, we improve the fieldâ□□s understanding of fibromyalgia, a pain condition linked with anxiety, depression, and subjective well-being, among other commonly comorbid traits and we make advances towards understanding shared genetic architecture of these related traits.

PrgmNr 3318 - Genetic architecture of risky sexual behavior: Correlational structure and causal effects

View session detail

Author Block: **N. Berley**[1], B. M. Huibregtse[2], L. M. Evans[2], M. C. Stallings[2]; [1]Virginia Commonwealth Univ., Richmond, VA, [2]Univ. of Colorado Boulder, Boulder, CO
Disclosure Block: **N. Berley:** None.
An ambiguous measure of risky sexual behavior (RSB), number of lifetime sexual partners, has been hypothesized to reflect either of two competing domains. The first is that it is related to reproductive fitness specifically success in the sexual selection of mates in males and the second is that it is a risk-taking behavior akin to speeding in an automobile or using alcohol and drugs. Using structural equation modelling to jointly estimate the genetic architecture of selected risk and fitness phenotypes, demonstrates that number of lifetime sexual partners is more closely related, genetically, to a latent risk factor as opposed to a fitness factor. Follow up results, from univariate and multivariate Mendelian randomization to assess casual effects using Egger regression were inconclusive. A two factor model indicates that number of sexual partners is best explained as a risk-related trait.

PrgmNr 3319 - Genome-wide association study of anthropometric traits in a health-disparate Hispanic population

View session detail

Author Block: **W. Zhu**[1], H. Polikowsky[1], H. M. Highland[2], M. Lee[3], J. B. McCormick[3], S. P. Fisher-Hoch[3], K. E. North[4], J. E. Below[5]; [1]Vanderbilt Univ., Nashville, TN, [2]Univ North Carolina at Chapel Hill, Chapel Hill, NC, [3]Univ. of Texas, Houston, TX, [4]Univ North Carolina, Chapel Hill, NC, [5]Vanderbilt Univ Med Ctr., Nashville, TN

Disclosure Block: **W. Zhu:** None.

Obesity is a common metabolic disorder associated with many chronic diseases, including type 2 diabetes, coronary heart disease, and certain types of cancer. The prevalence of obesity in the US adult population is ~42% and the prevalence is even higher, ~44.8%, among Mexican Americans. Low socioeconomic status is also often linked to obesity rates. Cameron County Hispanic Cohort (CCHC) is a population-based cohort established 17 years ago. It includes almost 5,000 individuals from low-income border communities, over 90% of whom are Mexican American. Hispanic/Latino populations are known to have higher rates of metabolic disease compared to non-Hispanic whites, and prior epidemiologic studies in CCHC have shown the obesity (51%) and diabetes (28%) rates are significantly higher than national estimates. This disproportionate burden of disease may make CCHC well-suited to study of obesity-related metabolic traits.

We performed a GWAS of waist-to-hip ratio in CCHC participants to identify for genetic factors associated with obesity. We removed pediatric measurements (age EX array and imputed to TOPMed phase 8 reference data following standard protocols. 2,946 individuals passed quality control and were included in our GWAS. Covariates including age, age squared, three principal components, and body mass index (BMI) were modeled using a generalized linear model to calculate residuals. Residuals were rank-based inverse normal transformed prior to analysis using EPACTS. Initial results identified five variants, chr8:6551122 and chr21:19511384 without BMI adjustment, and chr5:173508154, chr21:19511384, hr11:17563655 and chr21:20190723 without BMI adjustment. Additional GWAS on waist-to-height and height were also reported following the same analysis procedure.

PrgmNr 3320 - Genomic approaches to identify shared genetic architecture among comorbid phenomes of eye disease

View session detail

Author Block: **A. Scalici**[1,2], T. W. Miller-Fleming[1,2], D. Zhou[1,2], E. W. Knapik[1,2], N. J. Cox[1,2]; [1]Div. of Genetic Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr., Nashville, TN

Disclosure Block: **A. Scalici:** None.

Genome-wide association studies have identified a significant number of SNPs associated with human disease. However, many of these associations provide little insight into the underlying biological mechanisms or pathways involved in disease pathogenesis. Examining genetic pleiotropy is one means by which we can identify shared disease mechanism. Previous studies suggest that there is potential shared genetic architecture among different eye diseases. To better understand the potential shared underlying mechanisms of eye disease, we used both a gene-based approach to assess shared genetic architecture and a phenome-based approach to identify comorbidities and expand the nascent pathways associated with eye diseases. To assess shared genetic architecture among eye diseases, we used transcriptome-wide association study (TWAS) to test if imputed genetically regulated expression (GReX) in genotyped individuals of European ancestry (N=70,493) within BioVU is associated with eye disease status (Zhou et al., 2020). We identified two genes (*GPX7* & *AC016590.3*) that had significant associations to eye disease status. To identify the comorbid phenomes of eye disease, we conducted a phenome-wide association study (PheWAS) within non-genotyped subjects with at least three visits to VUMC in five years (N=685,820). Using phenomes significantly associated with eye disease case status, we calculated a phenotypic risk score (PheRS) by calculating a weighted sum of phenotypic effects, and applied this score to an independent population - genotyped subjects in BioVU (Bastarache et al., 2018). This PheRS based only on non-eye disease comorbid phenome is predictive of eye disease status. The PheRS was significantly associated with the predicted expression of six genes (*BBS5*, *NEU2*, *C4B*, *PAK1*, *RPL41*, *AC091100.1*), previously characterized as being involved in eye diseases. Functional analysis of these genes using zebrafish as a model system has the potential to shed light upon some of the common pathways and mechanisms of eye disease.

PrgmNr 3321 - Individual Imputation of Genetically Regulated Gene Expression in the Million Veteran Program

View session detail

Author Block: **S. Venkatesh**, W. Zhang, G. E. Hoffman, G. Voloudakis, P. Roussos; Icahn Sch. of Med. at Mount Sinai, New York, NY
Disclosure Block: **S. Venkatesh:** None.
Most neuropsychiatric disorders are moderately heritable but characterized by many genetic risk variants with weak effects. As such, it is difficult to point to direct causes or elucidate mechanisms of action. Despite the ease in gathering genetic data from humans, genetic data does not easily explain mechanistic effects. Gene expression on the other hand, can more easily explain mechanistic effects, but is harder to gather, especially in brain regions which are critical to the understanding of neuropsychiatric disease. To address this, we developed methods to impute genetically regulated gene expression (GREx) from genotypes and imputed GREx in over four hundred thirty thousand European individuals in the Million Veteran Program (MVP) for a wide variety of tissues and cell types. We use EpiXcan (based on PrediXcan) to develop machine learning models from training genotype, expression, and epigenetic data. We use custom scripts to impute individual GREx and perform a variety of downstream association analyses, including Principal Component Analysis (PCA), GREx Phenome Wide Association Studies (PheWAS), and Transcriptome Wide Association Studies (TWAS). Results show an overlap in Schizophrenia genes identified by individual level TWAS and those identified by summary level TWAS informed by GWAS. Inverse-variance meta-analyzed single gene imputation efforts across ancestries confirm clinical results obtained from COVID-19 positive individuals in both *IL10RB* and *IFNAR2*. GREx PheWAS for these particular genes using a novel negative binomial distribution for phecodes confirm COVID-19 related phenotypes. GREx presents a unique solution to integrate effects across the genome and increase sample size in gene expression analyses. We are pursuing the creation of additional EpiXcan models, improved statistical methods for downstream association analyses, and replication efforts across biobanks. We plan to perform these analyses in all ancestries, available EpiXcan and PrediXcan models, and phenotypes.

PrgmNr 3322 - Investigation of a complex phenotype involving craniosynostosis and cataracts leads to identification of a novel mutation related to Galactosemia type II

View session detail

Author Block: **R. H. Finnell**[1], B. Oliveira[2], T. Silva[2], y. lei[1], C. Souza[2], J. Filho[3], I. Schwartz[2]; [1]Baylor Coll. Med., Houston, TX, [2]Med. Genetics Service, Hosp. de ClÃnicas de Porto Alegre, Porto Alegre, Brazil, [3]Projeto Genomas Raros, Albert Einstein Hosp., SÃ£o Paulo, Spain
Disclosure Block: **R.H. Finnell:** None.
Introduction,Galactosemia is an inborn error of metabolism resulting in an accumulation of galactose metabolites, which can lead to cataracts and hepatic complications. Treatment consists of maintaining a lactose-restricted diet preferably initiated early in life. During a genetic investigation of a patient with a complex phenotype involving cataracts and craniosynostosis, a novel variant in the *GALK1* gene was identified. **Case report**, A 2yo male patient was referred for genetic evaluation due to bilateral dense cataracts. He is the fourth child of a non-consanguineous couple; his 18yo sister has a history of congenital cataract. Clinical examination revealed a metopic crest and trigonocephaly, with tomographic confirmation of craniosynostosis. Karyotype was normal and oligosaccharide chromatography suggested galactosemia but was excluded upon finding normal Galactose-1-phosphate-uridyltransferase activity. Initial dried-blood-spot galactose levels were 10 times above normal upper limits, which normalized after starting dietary treatment. His sibling with congenital cataracts had no clinical or biochemical evidence of galactosemia. He underwent lensectomy and surgical correction of craniosynostosis. WGS led to identification of a large novel homozygous truncating duplication in *GALK1* gene affecting splice site c.166-16_227dup (NM_000154.2). **Discussion**:There are previous reports of the association between craniosynostosis and classic galactosemia in rare cases of homozygous contiguous deletion affecting *GALT* and *IL11RA* genes (9p13.3); however, there exists no data on craniosynostosis with other forms of galactosemia. The identified novel likely pathogenic variant explains both galactosemia and cataract phenotypes. No pathogenic variants related to craniosynostosis phenotype were identified. **Conclusion**, This case report describes a novel variant related to galactosemia type II. Future investigations involve analyses at the transcriptional level.

PrgmNr 3323 - Modeling the longitudinal changes of ancestry diversity in the MVP cohort: systematic comparison of methods

View session detail

Author Block: **F. Wendt**[1,2], G. A. Pathak[2,1], J. Vahey[3,4], D. Koller[1,2], X. Qin[3,4], K. Harrington[5], D. F. Levey[1,2], F. De Angelis[1,2], A. De Lillo[1], B. Cabrera-Mendoza[1,2], L. M. Duong[2], J. Gelernter[1,2], M. Aslan[6], D. Provenzale[4,3], D. Helmer[7], E. R. Hauser[3,4], R. Polimanti[1,2]; [1]Yale Sch. of Med., New Haven, CT, [2]VA CT Hlth.care System, West Haven, CT, [3]Duke Univ., Durham, NC, [4]Durham VA Med. Ctr., Durham, NC, [5]Boston Univ. Sch. of Med., Boston, MA, [6]VA Clinical Epidemiology Res. Ctr. (CERC), VA Connecticut Hlthcare System, West Haven, CT, [7]Michael E. DeBakey VA Med. Ctr., Houston, TX

Disclosure Block: **F. Wendt:** None.

Background/Aims: The birth years of MVP participants span nearly 100 years, but it remains unclear (i) if longitudinal changes in ancestry are present and (ii) how these changes can be accounted for in the MVP cohort to conduct reliable gene-phenotype association analyses.

Methods: We divided MVP participants into five birth cohorts (BCs) of approximately 125,000 individuals each and compared ancestry classifications using HARE (harmonized ancestry and self-reported race/ethnicity) and a random-forest genetic principal component clustering using the 1KGP and the HGDP reference panels. We used these ancestry classifications in genome-wide association studies (GWAS) of height, a highly polygenic trait with strong potential for population-stratification bias. GWAS were conducted in each BC stratified by ancestry, including age, sex, and ten within-ancestry principal components. Linkage Disequilibrium Score Regression (LDSC) was used to calculate the intercepts and attenuation ratios to assess residual population stratification in the GWAS statistics. These metrics were then compared with two-sided Z-tests.

Results: Relative to the oldest BC, more contemporary HARE Europeans (EUR; ~0.2% reduction in EUR), Africans (AFR; ~2% reduction in EUR), and Hispanics (HIS; ~3% reduction in EUR) had a significant reduction in EUR ancestry proportion. Conversely, HARE East Asians (EAS) showed an increase in EUR ancestry proportions over time. The most contemporary EAS BC had three times greater EUR ancestry than the oldest BC (pConclusions: We provide the first characterization of MVP cohort ancestry diversity over time, highlight potential biases in gene discovery, and show how the methods available can perform differently in accounting adequately for human genetic variation.

PrgmNr 3324 - Novel Osteoarthritis Genetic Variants Identified using Multi-Ancestry Analyses from 473,388 Participants in the MVP and the UK Biobank

View session detail

Author Block: **M-L. N. McDonald**[1,2], P. Lakshman Kumar[1,2], V. Srinivasasainagendra[1,2], J. Richman[1,2], S. Pinson[1], A. Rocco[1], A. Nair[1], R. Dennis[3], V. Jagadale[3], C. Brown[1], H. K. Tiwari[4], M. Bamman[1,5], J. Singh[1,2]; [1]Univ. of Alabama at Birmingham, Birmingham, AL, [2]Birmingham VA Hlth.care System (BVAHS), Birmingham, AL, [3]Central Arkansas Veterans Hlth.care System (CAVHS), Little Rock, AR, [4]Univ Alabama at Birmingham, Birmingham, IN, [5]Birmingham VA Hlth.care System (BVAHS),, Birmingham, AL

Disclosure Block: **M.N. McDonald:** None.

Osteoarthritis (OA) is a progressive, degenerative joint disease with a poorly understood etiology primarily limited in clinical care to symptom management related to pain and inflammation. The prevalence of OA is high at 40% and associated with great disability. To date there have been no large multi ancestry genetic studies of OA. We leveraged the unique resources of 473,388 participants in the combined Million Veteran Program (MVP) and UK Biobank to address this gap including participants of European, African, Asian and Hispanic descent. In doing so, we discovered OA associated genetic variants in or near 71 genes and replicated findings from 52 previously identified OA variants or genes. We also discovered evidence some OA associated regions may be robust to population ancestry. Our novel findings provide insight to how strategies for managing metabolic syndrome may be used to help prevent and manage OA.

PrgmNr 3325 - Novel phenotyping approaches: Empirical Bayes methods improve heritability of dietary phenotypes from repeat diet questionnaire data

View session detail

Author Block: **J. B. Cole**[1], J. C. Florez[2], J. N. Hirschhorn[3]; [1]The Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Gen Hosp, Boston, MA, [3]Boston Children's Hosp., Boston, MA
Disclosure Block: **J.B. Cole:** None.
Unhealthful dietary intake is a leading risk factor for metabolic disease and mortality, yet the precise aspects of diet which influence health are largely unknown. Recent advent of large biobanks has enabled genomic discovery for modestly heritable traits, such as dietary intake. However, diet represents a complex set of multifactorial, dynamic, and highly correlated variables, and therefore studies on the genetics of diet would greatly benefit from novel phenotyping approaches. Previously we identified loci unique to either univariate or multivariate diet traits derived from the same food frequency data in UK Biobank (UKB).

Here, we aim to extract meaningful biological traits from 24-hour diet recalls, consisting of hundreds of questions about the specific foods/drinks and quantities consumed in the prior 24 hours. While an individualâ□□s diet on any given day is highly variable, many individuals in UKB took this questionnaire multiple times, providing an opportunity to apply methods to incorporate certainty from repeated measures. In a population of 176K individuals of European ancestry in UKB, for which 47K individuals took the questionnaire >2 times and 2K individuals took it all 5 times, we applied empirical Bayes methods to weight an individualâ□□s food values by the number of times they took the questionnaire. Specifically, we applied this to how often an individual ate a specific food as a proportion of the number of questionnaires completed and to how much an individual ate using weighted averages. We used LD score regression heritability analysis to benchmark Empirical Bayes to their non-Bayes counterparts: an ever/never eaten phenotype and a simple average for continuous variables, respectively.

The heritability estimates for 139/158 (89%) Empirical Bayes food proportions were larger than their ever/never counter-part phenotypes, with the largest difference in fruit intake (heritability from 2 to 5%) and the heritability estimates for 206/242 (85%) Empirical Bayes weighted averages were larger than simple averages, particularly for less commonly eaten foods where one data point is not an accurate reflection of eating habits. Interestingly, when we weight averages for derived nutrients, we find no evidence for improvement. This is in line with epidemiological work demonstrating there is larger day-to-day variation in food consumption vs. nutrients.

We demonstrate that novel phenotyping approaches, and in particular Empirical Bayes methods for repeated data, can improve genomic discovery. These methods can be extended to many other complex traits with repeated measures, with particular utility for traits with high intra-individual variation.

PrgmNr 3327 - Programming biobank-scale relatedness analysis improvements

View session detail

Author Block: **G. Evans**[1], H-H. Chen[2], C. D. Huff[3], R. J. Bohlender[4], J. E. Below[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ Med Ctr., Nashville, TN, [3]U.T. MD Anderson Cancer Ctr, Houston, TX, [4]Univ. of Texas MD Anderson Cancer Ctr., Houston, TX
Disclosure Block: **G. Evans:** None.
The program PRIMUS (Pedigree Reconstruction & Identification of the Maximum Unrelated Set) was developed to facilitate rapid pedigree reconstruction using genome-wide estimates of identity-by-descent (IBD). However, PRIMUS is sensitive to IBD estimation bias and fails to assess relationship complexity prior to attempting reconstruction, both of which can cause execution errors. We are in the process of completing a total overhaul of the original PRIMUS codebase to upgrade it both computationally and analytically as well as integrate it with relatedness estimation software ERSA (Estimation of Recent Shared Ancestry) and PADRE (Pedigree-Aware Distant Relatedness Estimation) to form COMPADRE (Combined PADRE) to streamline the relationship detection and pedigree construction pipeline.

We have accomplished several upgrades thus far in the COMPADRE development undertaking. First, we implemented a quality control framework that ingests family network data (non-directional relatedness graphs) at scale to evaluate the likelihood of successful downstream pedigree reconstruction given network size and sparsity. This framework improved the proportion of completed pedigree reconstructions by ~25% in a large data setting (n~100,000 from Vanderbilt University Medical Centerâ□□s biobank, BioVU). We also successfully introduced a series of multithreading-based computational improvements to the overall PRIMUS infrastructure, which improved overall runtime by ~35% and lowered computational footprint by ~20% particularly during execution of pedigree reconstruction. Finally, we are currently improving PRIMUSâ□□s relationship likelihood assessment by harnessing ERSA output data to introduce a composite probability vector that increases pedigree confidence and reduces runtime errors by clarifying ambiguous second- and third-degree relationship estimation. This, and other analytically novel components of the PRIMUS rewrite are being tested using an augmented data simulation pipeline adapted from past collaborators that is being directly incorporated for enhanced user experience.

Our findings thus far demonstrate that targeted computational and analytical updates can successfully reduce bias and improve pedigree reconstruction in PRIMUS. Future work will continue to introduce and refine novel methods to proactively assess COMPADREâ□□s ability to dynamically analyze biobank-scale data. COMPADRE development is also happening alongside that of a shared segments repository, specifically engineered to maximize relatedness analysis through intimate connections to both large DNA biobanks and the tools needed (such as PRIMUS) for analysis.

PrgmNr 3328 - Reanalysis of exome data via Diagnosticator identifies significant variant reclassification with updated ClinGen SVI-WG recommendations

View session detail

Author Block: **E. Cocchi**[1], M. Elliott[2], N. Vena[3], A. G. Gharavi[4]; [1]Columbia Univ., New York, NY, [2]Columbia Univ. Irving Med. Ctr., Calgary, AB, Canada, [3]Hammer Hlth.Sci. Building, New York, NY, [4]Columbia Univ, New York, NY

Disclosure Block: **E. Cocchi:** Grant/Contracted Research Support (External); American Society of Nephrology.

The current standard in sequence variant interpretation is the 2015 ACMG guidelines, which have improved reproducibility and provided a common nomenclature. To further improve consistency of reported results, ClinGenâ□□s Sequence Variant Interpretation Working Group (SVI WG) has developed recommendations that clarify specific ACMG criteria, however the impact of the revised criteria on variant reclassification is unknown.

We previously evaluated 2144 Columbia patients with kidney disease who underwent exome sequencing (ES) using the ACMG 2015 guideline. We identified 193 pathogenic or likely-pathogenic kidney-related variants, with a diagnostic yield of 9%. We used the *Diagnosticator* clinical annotation tool to reanalyze these cases incorporating both newly published genetic information and the SVI WG criteria recommendations. *Diagnosticator* is a flexible annotation tool which can rapidly implements the SVI WG recommendations and is faster and more accurate than manual curation.

Reanalysis of the entire cohort, incorporating newly available genetic data and the recent SVI WG recommendations, led to a change in the ACMG classification in 79 (40.9%) of the variants. 53 (27.4%) variants downgraded their overall classification: 18 (9.3%) variants changed from P to LP, 33 (17.0%) variants changed from LP to VUS, and 2 (2.5%) variants changed from P to VUS. 26 (13.5%) variants were upgraded from LP to P. These changes are driven by criteria with SVI WG recommendations (PVS1, PM2, PM3, PP5) and with new information available (PS1, PS4, PM3, PM5, PP1, PP3).

We demonstrate the utility of *Diagnosticator* as a tool for rapid reanalysis of ES data, and show that incorporating the ClinGen SVI WG recommendations and newly available genetic data lead to reclassification for a significant number of variants. These data motivate reanalysis of ES data at standardized intervals to assess impact on clinical care.

PrgmNr 3329 - Study of over one million people finds that novel and previously reported familial genetic effects impact stuttering risk at a population-level

View session detail

Author Block: **H. Polikowsky**[1], D. M. Shaw[2], L. E. Petty[2], H-H. Chen[1], Y. Yu[3], 23andMe Research Team, C. D. Huff[3], N. J. Cox[2], S. J. Kraft[4], J. E. Below[2]; [1]Vanderbilt Genetics Inst., Vanderbilt Univ Med Ctr., Nashville, TN, [2]Vanderbilt Univ Med Ctr., Nashville, TN, [3]U.T. MD Anderson Cancer Ctr., Houston, TX, [4]Behavioral Speech & Genetics Lab, Wayne State Univ., Detroit, MI
Disclosure Block: **H. Polikowsky:** None.
Despite an ~1% prevalence of persistent stuttering in the general adult population, developmental stuttering, characterized by prolongation of sounds and interruptions in speech, remains a largely idiopathic disfluency disorder. Its complex mode of inheritance combined with thus-far underpowered genetic studies contribute to the challenge of identifying and reproducing genes implicated in developmental stuttering susceptibility. Here, we report the results of a genome-wide association study of stuttering for >1 million individuals. We conducted a GWAS in individuals genotyped by 23andMe, Inc. who consented to participation. Cases were defined as those who self-reported ever having a stammer or a stutter (N=99,776), with all other individuals considered controls, comprising a total of 1,023,243 multi-ethnic participants. We performed ancestry and sex-specific logistic regression models assuming an additive model for allelic effects and including age, five principal components, and genotype platforms as covariates: 18 genome-wide significant hits were identified. The most significant risk variant (p = 5.8 x 10-16) was identified in the European-female (rs13107325, RAF= 8.16%, OR = 1.117) GWAS for a missense variant in *SLC39A8*, a gene encoding solute-carrier proteins critical for cellular transport in mitochondria. The European-male GWAS also found a strong signal at this locus. Interestingly, in the European-male GWAS we identified a near-significant (p = 9.8 x 10-7) risk variant (rs111790048, RAF= 0.005%, OR = 10.854) in the intronic region of *GNPTG*, a gene previously implicated in a linkage study of developmental stuttering by Kang et. al. (NEJM, 2010). To evaluate our hypothesize that developmental stuttering represents a complex polygenic trait akin to neurogenic traits, we developed a preliminary polygenic-risk model using variants from the European male-specific GWAS. We tested this PRS model within an independent European-male clinically-ascertained developmental stuttering GWAS and achieved moderate model performance (AUC= 0.6134), suggesting that our self-report stuttering phenotype is capturing genetic effects for clinically-confirmed cases. Independent replication analysis in this clinically-ascertained GWAS is ongoing. These exciting preliminary results represent the first report of genetic contributions to developmental stuttering at a population-level. Our results highlight the polygenicity of the trait and suggest that variation at both novel and previously reported loci identified via family-based analyses impact population-level stuttering risk.

PrgmNr 3330 - The contribution of common versus low-frequency variation to hearing loss heritability

View session detail

Author Block: **L. Dobbyn**, K. Praveen, L. Gurski, A. Ayer, M. Drummond, A. Moscati, M. Ferreira, G. Coppola, E. Stahl, DiscovEHR, Regeneron Genetics Center; Regeneron Pharmaceuticals, Tarrytown, NY
Disclosure Block: **L. Dobbyn:** Salary/Employment; Regeneron Genetics Center.
INTRODUCTION: Age-related hearing loss is prevalent and heritable with an established common variant component to its genetic architecture. Genome-wide association studies (GWAS) have identified over 40 associated common variant loci and shown significant SNP heritability with estimates ranging from ~0.08 to ~0.19. We previously conducted a combined GWAS+ExWAS of hearing loss in 115,675 and 344,085 European-ancestry cases and controls across four cohorts (Geisinger DiscovEHR, UK Biobank, Mount Sinai *BioMe* Biobank, and the MalmÃ¶ Diet and Cancer Study), resulting in 53 common and 41 rare variant associations. Our analysis not only confirmed this common variant component but also implicated rare variation in the etiology of age-related hearing loss. To quantify the heritability due to variants of lower frequencies and to characterize the relative contributions from rare and common variation, we conducted heritability analyses using an approach (Gazal et. al. 2018) that extends stratified LD score regression (S-LDSC) to low-frequency variation. METHODS: Through multivariate regression of association statistics on LD scores from our hearing loss meta-analysis, we estimated heritability for 14 functional categories. To capture both common and rare variation in our LD scores, we generated scores using a reference panel created by merging imputed and exome data for 6,189 European-ancestry samples from the UK Biobank. We calculated LD for seven functional categories (coding-synonymous, coding-nonsynonymous, 5-prime-UTR, 3-prime-UTR, splice site, intronic, and intergenic) each split into a common (MAF>=0.05) and low-frequency (MAF0.001, and variants identified by machine learning as false-positives on the basis of sequence QC metrics were excluded. Association statistics were filtered at MAF>0.001 and INFO>0.9. RESULTS: Total liability-scale heritability (h^2_{total}) was estimated to be 0.076, with a common variant heritability (h^2_{CV}) of 0.071 and a low-frequency variant heritability (h^2_{LFV}) of 0.005. Common intronic variants were significantly (Bonferroni P2, and low-frequency intronic and intergenic variants were depleted for h^2. Categories with the largest enrichments were common coding-synonymous, 5-prime-UTR, and coding-nonsynonymous (enrichments of 52.5x, 39.6x, and 18.4x, respectively), and low-frequency 5-prime-UTR and coding-nonsynonymous (enrichments of 16.9x and 11.7x). Our results show that despite having a small absolute contribution, low-frequency variation is significantly enriched in hearing loss heritability.

PrgmNr 3331 - Transcriptome-wide association study and eQTL colocalization identify potentially causal genes responsible for bone mineral density GWAS associations

View session detail

Author Block: **B. M. Al-Barghouthi**[1], W. Rosenow[1], K-P. Du[1], J. Heo[1], R. Maynard[2], L. D. Mesner[1], G. Calabrese[1], A. Nakasone[3], B. Senwar[4], L. Gerstenfeld[5], V. Ferguson[4], C. Ackert-Bicknell[2], E. Morgan[3], D. L. Brautigan[1], C. R. Farber[6]; [1]Univ. of Virginia, Charlottesville, VA, [2]Univ. of Colorado Anschutz, Aurora, CO, [3]Boston Univ., Boston, MA, [4]Univ. of Colorado, Boulder, Boulder, CO, [5]Boston Univ. Med. Ctr., Boston, MA, [6]Univ Virginia, Charlottesville, VA

Disclosure Block: **B.M. Al-Barghouthi:** None.

Genome-wide association studies (GWASs) for bone mineral density (BMD) have identified over 1,100 associations to date. However, identifying causal genes implicated by such studies has been challenging. Recent advances in the development of transcriptome reference datasets and computational approaches such as transcriptome-wide association studies (TWASs) and expression quantitative trait loci (eQTL) colocalization have proven to be informative in identifying putatively causal genes underlying GWAS associations. Here, we used TWAS/eQTL colocalization in conjunction with transcriptomic data from the Genotype-Tissue Expression (GTEx) project to identify potentially causal genes for the largest BMD GWAS performed to date. Using this approach, we identified 512 genes as significant (Bonferroni PPP6R3, the gene with the strongest support from our analysis which was not previously implicated in the regulation of BMD, for further investigation. We observed that *Ppp6r3* deletion in mice decreased BMD. In this work, we provide an updated resource of putatively causal BMD genes and demonstrate that *PPP6R3* is a putatively causal BMD GWAS gene. These data increase our understanding of the genetics of BMD and provide further evidence for the utility of combined TWAS/colocalization approaches in untangling the genetics of complex traits.

PrgmNr 3332 - Whole-exome sequencing of a family with specific language impairment suggests *BUD13* as a novel SLI candidate gene

View session detail

Author Block: **E. M. Andres**, K. K. Earnest, C. Zhong, M. L. Rice, M. H. Raza; Univ. of Kansas, Lawrence, KS
Disclosure Block: **E.M. Andres:** None.
Specific language impairment (SLI) is characterized by a delay in the mastery of language and family studies have shown aggregation of cases in medium, large and extended families. Genetic investigation of SLI is complicated by behavioral heterogeneity of the disorder. Several candidate genes have been suggested but the biological mechanisms underlying the transmission and expression of SLI are still not well understood. The current study utilized a family-based approach, reducing behavioral trait and genotypic variance, with the aim of identifying co-segregating rare exonic variants. Whole-exome sequencing was performed in six individuals of a family with SLI (*N*=11). Three co-segregating rare non-synonymous variants (MAF â¤ 0.005) were identified in three genes: *BUD13*, *APLP2* and *NDRG2*, suggesting oligogenic inheritance of SLI in this family. Then, we Sanger sequenced all coding exons (including flanking intronic regions) of the identified genes in additional unrelated individuals with SLI (*N*=175). We observed variants in both BUD13 and APLP2 (in the 3â□□UTRs) in a second family, indicating further evidence of an oligogenic basis of SLI. In total, 13 additional rare non-synonymous or 3â□□UTR point mutations were observed across the three identified genes in 18 unrelated individuals with SLI. Nine of the variants were observed in *BUD13*, a rate that reached genome-wide significance according to comparison with the rate of similar variants in the global population within the 1000 Genomes database. Of the rare variants identified in *BUD13*, three were observed in more than one unrelated individual with SLI. Bud13 is a component of the retention and splicing (RES) complex with an important role in controlling gene expression during cellular differentiation. Moreover, a previous animal model has shown Bud13 has a role in neural phenotypes. The evidence together suggests *BUD13* is a novel candidate for SLI.

PrgmNr 3333 - Year of birth and polygenic scores in the Electronic Health Records

View session detail

Author Block: **M. Niarchou**[1], D. Zhou[1], R. Kember[2], R. Karlsson LinnÃ©r[3], G. Voloudakis[4], L. Yengo[5,6], G. Chen[7], C. Chabris[8], L. K. Davis[9]; [1]Vanderbilt Univ., Nashville, TN, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Vrije Univeristeit Amsterdam, Amsterdam, Netherlands, [4]Icahn Sch. of Med., New York, NY, [5]Brisbane, Australia, [6]Inst. for Molecular BioSci., University of Queensland, Australia, [7]Univ. of Wisconsin-Madison, Wisconsin, WI, [8]Geisinger Hlth.System, Lewisburg, PA, [9]Vanderbilt Univ Med Ctr., Nashville, TN

Disclosure Block: **M. Niarchou:** None.

Polygenic scores (PGS) are already been used in the clinical context and they may eventually be implemented in routine clinical practice as clinical biomarkers. It is important to have a better understanding of how factors not directly related to disease biology may affect the value of the PGS before applying them in clinical settings. For example, we and others have observed a correlation between Year of Birth (YoB) and PGS for different health conditions and quantitative traits. We further investigated this phenomenon in the Vanderbilt Biobank (BioVU) and replicated the finding in biobanks in the PsycheMERGE network. We first generated PGS trained on summary statistics from the latest genome-wide association studies of a variety of health conditions (i.e., ADHD, coronary artery disease (CAD), major depressive disorder (MDD), insomnia, Alzheimerâ□□s disease (AD)), quantitative traits (BMI, height, educational attainment), and blood traits (i.e., white blood cell count (WBC), red blood cell count (RBC), red blood cell count distribution width (RDW)), and applied them to genome-wide data from 66,904 unrelated individuals of European Ancestry. We then fitted separate linear regression models to test whether each PGS (predictor variable) was associated with YoB (outcome variable) after adjusting for sex and 10 principal components of ancestry. All PGS, except for scores trained on blood cell traits, were significant predictors of YoB. Next, we performed ten different sets of analyses to test four possible hypotheses for the observed correlation. We hypothesized that the correlations may be induced by 1) the effects of age in the discovery GWAS; 2) differences in population structure by birth cohort; 3) increase in relatedness across birth cohorts; or 4) changing ascertainment bias within the medical system over time. There was no single factor that stood out as accounting for the trends, and the observed relationship between YoB and PGS was robust to most adjustments. Overall, our findings indicate that the correlation between birth year and PGS is likely due to a combination of factors and the extent to which they are implicated in the associations vary depending on the trait.

PrgmNr 3334 - A global transcriptomic approach reveals sex-specific differences in a middle-aged frailty cohort

View session detail

Author Block: **N. L. Pacheco**, N. Noren Hooten, Y. Zhang, C. S. Prince, N. A. Mode, N. Ezike, K. G. Becker, A. B. Zonderman, M. K. Evans; Natl. Inst. on Aging, NIH, Baltimore, MD
Disclosure Block: **N.L. Pacheco:** None.
Frailty is a clinical syndrome described as reduced physiological reserve and increased stress or vulnerability. Usually studied in older groups, recent work shows frailty occurs in middle-aged individuals and is associated with increased mortality. Identifying biologic pathways underlying frailty development in middle-aged adults is crucial to detect and prevent frailty as well as reduce premature mortality. Previous examination of global gene expression changes in a middle-aged cohort of African Americans and whites from the Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) study demonstrated that activation of inflammatory genes and pathways was significantly altered by frailty status and race. As there are well-established sex differences in frailty burden and mortality, it is important to decipher transcriptome-wide differences in frailty-associated genes by sex. In this study, we sought to identify novel genes and pathways associated with sex and frailty in a diverse middle-aged cohort using RNA-Sequencing. Differential gene expression and pathway analyses were performed for 4 comparison groups: 1) frail females (FRAF, n = 4) vs non-frail females (NORF, n = 4); 2) frail males (FRAM, n = 4) vs non-frail males (NORM, n = 4); 3) FRAM vs FRAF; and 4) NORM vs NORF as a control. We evaluated exclusive significant genes and pathways, as well as overlaps, between the 4 comparison groups. Over 80% of the significant genes exclusive to FRAF vs NORF, FRAM vs NORM, and FRAM vs FRAF, respectively, were novel and associated with inflammatory and metabolic pathways. Pathways exclusive to FRAF vs NORF were associated with inhibition of inflammatory pathways. FRAM vs NORM exclusive pathways were associated with decreased musculoskeletal protein turnover. Pathways exclusive to FRAM vs FRAF were associated with activated metabolite degradation. Notably, genes with key inflammasome and apoptotic roles in the Coronavirus Pathogenesis Pathway were activated in FRAM vs FRAF. We also confirmed frailty-associated changes in gene expression from other previous studies for *OTUD1* (FRAF vs NORF), *LRG1* (FRAM vs NORM), and *TNF* (FRAM vs FRAF), as well as pathways like IL-1 signaling (FRAF vs NORF), B cell development (FRAM vs NORM), and Salvage Pathways of Pyrimidine Deoxyribonucleotides (FRAM vs FRAF). Our results indicate sex-specific transcriptional changes occur in middle-aged frailty, advancing knowledge on frailty progression and identification of potential therapeutic targets to prevent frailty.

PrgmNr 3335 - A single cell chromatin accessible and transcriptome atlas of the human heart facilitates identification of risk variants and genes of Atrial Fibrillation

View session detail

Author Block: **S. Pott**[1], A. Selewa[2,1], M. Wasney[1], L. Smith[1], H. Eckart[1], I. Moskowitz[1], A. Basu[1], X. He[3]; [1]Univ. of Chicago, Chicago, IL, [2]Chicago, IL, [3]Univ Chicago, Chicago, IL
Disclosure Block: **S. Pott:** None.
Genome-wide association studies (GWAS) have identified thousands of loci for hundreds of diseases. However, most trait-associated variants are located in noncoding regions with no clear functions, making it difficult to identify causal variants, their targeted genes, and the specific cell type(s) affected. Single-cell genomics promises to close these gaps, by de-convoluting complex cellular mixtures into constituent cell types, and annotating functional elements in each cell type. In this work, we combined single cell RNA-seq and ATAC-seq to profile the human heart. We detected 8 major cell types across both modalities and identified 349,000 open chromatin regions (OCRs), half of which are cell-type-specific. Leveraging this dataset, we identified lineage-specific transcription factors (TFs), and assigned putative target genes to 39% of distal OCRs.
We used this dataset to study the genetics of heart-related traits. Risk variants of Atrial Fibrillation (AF) are >10-fold enriched in cardiomyocytes (CMs) but not other cell types. Taking advantage of this enrichment pattern, we used a Bayesian statistical framework to fine-map causal variants of AF, favoring variants in CM-OCRs. This leads to 54 variants at posterior inclusion probability (PIP) > 0.5 - suggesting >50% probability of being causal variants, nearly doubling the number of variants under fine-mapping without functional information. This list of SNPs show a high proportion of overlapping heart H3K27ac regions, and TF targets from ChIP-seq. Furthermore, we developed a novel computational procedure that aggregates all putative causal variants and combines multiple sources of information linking SNPs to genes, to identify putative causal genes. This procedure identified 53 genes at 'gene-level PIP' > 0.5. A large fraction of these genes are TFs, ion channels, and cardiac signaling proteins with known roles in heart development and/or AF. Our results support a model where a gene regulatory network centered on TFs, including TBX5, GATA4, NKX2-5, and PITX2, underlies the AF genetic risk. Together, our study provides a comprehensive map of AF risk variants and genes, and demonstrates the power of combining single-cell genomics and advanced fine-mapping to reveal the genetic basis of complex traits.

PrgmNr 3336 - Blood-based differential DNA methylation and gene expression at the *TREM* locus in mild traumatic brain injury and Alzheimerâ□□s disease

View session detail

Author Block: **L. M. Bekris**[1], G. E. Weber[1], E. D. Tuason[1], M. Khrestian[1], S. Formica[1], Y. Shao[1], S. Rao[2], J. Pillai[2], J. B. Leverenz[2]; [1]Cleveland Clinic Lerner Res. Inst., Cleveland, OH, [2]Cleveland Clinic Lou Ruvo Ctr. for Brain Hlth., Cleveland, OH
Disclosure Block: **L.M. Bekris:** None.
Alzheimerâ□□s disease is a devastating neurodegenerative disorder and the most prevalent form of dementia worldwide. Alzheimerâ□□s disease (AD) has long been characterized by the presence of amyloid beta (AÎ²) plaques and tau neurofibrillary tangles in the brain, which makes prediction difficult and diagnosis definitive only at postmortem analysis. Triggering receptor expressed on myeloid cells 2 (TREM2) regulates inflammation in the periphery and brain. TREM2 has been implicated in AD due to its affinity for AD-related apolipoprotein E and binding of AÎ² to increase its uptake; mutations in *TREM2* have been shown to increase risk for AD and other neurodegenerative disorders. *TREM2* is part of a complex DNA locus that also includes other TREM and TREM-like (TREML) proteins. Inflammation can affect the epigenetic profile of circulating cells such as DNA methylation changes, resulting in altered expression of proteins. RNA sequencing of TREM and TREML transcripts provides insight into the molecular effects of DNA methylation. One such inducer of systemic acute and chronic inflammation that has been linked to AD is traumatic brain injury. To our knowledge this is the first study to compare methylation levels of the entire *TREML1-TREM1* locus in peripheral blood leukocytes in people with a history of TBI compared to AD, and correlate those findings to levels of RNA transcripts. It is unknown whether other regulatory region CpG sites in the *TREM* locus are differentially methylated in AD, and how this relates to levels of sTREM2 and AD biomarkers. Our previous findings indicate that plasma sTREM2 and cytokines are elevated in the mild cognitive impairment stage of AD, compared to later stage AD, suggesting an initial alteration of the immune system that is dampened in the later stages of AD. Therefore, we set out to determine whether *TREM* locus DNA methylation underlies this alteration in sTREM2 level. We hypothesized that the *TREM* locus would be differentially methylated in our cohort of people with TBI and AD compared to age- matched, cognitively normal controls, and correlate with known AD biomarkers. We found that genomic regions across the *TREM* locus exhibited differential methylation where in general promoters are hypomethylated while non-promoter regions are hypermethylated. Our preliminary findings also show *TREM*-locus methylation and transcriptome profiles differ between active and retired professional fighters and older adults with AD compared to age-matched cognitively normal controls. We anticipate that future studies will contribute to a better understanding of how *TREM2* is regulated during disease progression and whether it is a feasible therapeutic target.

PrgmNr 3337 - Comparing gene expression imputation resources in a diverse Latin American Parkinsonâ□□s disease cohort

View session detail

Author Block: **J. French-Kwawu**[1], V. Borda[2], D. Loesch[3], T. A. Thornton[4], I. Mata[5], T. D. O'Connor[1], on behalf of LARGE-PD; [1]Inst. of Genome Sci., Univ. of Maryland Baltimore, Baltimore, MD, [2]Univ. of Maryland Sch. of Med., Petropolis, Brazil, [3]Univ. of Maryland Baltimore, Baltimore, MD, [4]Univ of Washington, Seattle, WA, [5]Lerner Res. Inst., Genomic Med., Cleveland Clinic Fndn.,, Cleveland, OH
Disclosure Block: **J. French-Kwawu:** None.
Background: Genome wide association studies (GWAS) have proven to be a useful tool to identify variants associated with disease but interpreting results can be challenging. Imputing gene expression for GWAS results can identify genes involved in disease and highlight pathways involved in disease progression. However, many prediction models for gene expression are based on individuals of European ancestry. We compare imputed gene expression in the Latin American Research Consortium on the Genetics of Parkinsonâ□□s Disease (LARGE-PD) using prediction models from two cohorts - one of Latin American individuals and one of predominantly European ancestry individuals with a diversity of tissues (GTEx). Methods: We imputed gene expression for 1504 individuals from LARGE-PD using PrediXcan with four elastic net prediction models: three tissues from GTEx (whole blood, basal ganglia, and substantia nigra) and whole blood from Latin American individuals within the Multi-Ethnic Study of Atherosclerosis (MESA) cohort. We estimated the correlations of gene-level expression imputation for each pair of the four models. Results: The correlation of gene expression levels in GTEx basal ganglia and substantia nigra is 0.678 (pFunding: R01 NS112499 and T32 AG00262

PrgmNr 3338 - Cross-omics analysis of DNA methylation and gene expression in post-mortem 41 brain tissues of opioid use disorder patients and controls

View session detail

Author Block: **A. Liu**[1], Y. Dai[2], E. Mendez[3], R. Hu[4], G. Fries[5,3], K. Najera[3], S. Jiang[6], T. Meyer[3], L. Stertz[3], P. Jia[4], C. Walss-Bass[7], Z. Zhao[2,3]; [1]Dept. of Epidemiology, Human Genetics and Environmental Sci., Sch. of Publ. Hlth., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [3]Louis A. Faillace, MD, Dept. of Psychiatry and Behavioral Sci., McGovern Med. Sch., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [4]Ctr. for Precision Hlth., Sch. of BioMed. Informatics, The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, Houston, TX, United States, Houston, TX, [5]Ctr. for Precision Hlth., Sch. of BioMed. Informatics, The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [6]Ctr. for Precision Hlth., Sch. of BioMed. Informatics, The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, Houston, TX, [7]Louis A. Faillace, MD, Dept. of Psychiatry and Behavioral Sci., McGovern Med. Sch., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, United States, Houston, TX

Disclosure Block: **A. Liu:** None.

Background: Opioid use disorder (OUD) affects millions of people, causing nearly fifty thousand deaths annually in the United States. While opioid exposure and OUD are known to cause widespread transcriptomic and epigenetic changes, few studies have been conducted in human brain samples. Moreover, understanding how OUD affects the brain at multi-omics scale could help to decipher disease pathogenesis and shed light on OUD treatment. **Methods:** To fill in the gap, we generated genome-wide transcriptomic and DNA methylation profiles from post-mortem brain samples of 22 OUD subjects and 19 non-psychiatric controls. After performing quality control and normalization, we applied weighted gene co-expression network analysis (WGCNA) to identify network modules and genetic markers consistently associated with OUD at transcriptomic and methylomic levels. We further employed cross-omics analyses at the network module and gene levels to uncover OUD-specific regulatory networks. We performed hypergeometric tests at the network module level to assess overrepresented genes in OUD-associated co-methylation and co-expression modules. Functional enrichment analyses were performed on identified genes for biological interpretation. At the gene level, we examined global correlations between DNA methylation and gene expression changes. IngenuityÂ® Pathway Analysis (IPA) was used to identify canonical pathways and gene regulatory relationships underlying OUD pathogenesis for the genes with high methylation and gene expression variations. **Results:** In WGCNA, we found six OUD-associated co-expression gene modules and six co-methylation modules (p_{FDR} FDR = 1.16Ã□10^{-8}), gliogenesis (GO:0042063, p_{FDR} = 1.86Ã□10^{-8}), and astrocyte differentiation (GO:0048708, p_{FDR} = 2.26Ã□10^{-8}). At the gene level cross-omics analysis, we found a large proportion of the genes followed standard DNA methylation-gene expression regulatory patterns. Among the few exceptions, we identified their upstream transcription regulators that might bypass the regulations. **Conclusion:** Our integrative analysis of multi-omics data in OUD post-mortem brain samples suggested complex gene regulatory mechanisms involved in gene expression, transcription regulators, and epigenetic modification. Taken together, the current findings suggested that regulation of astrocyte and glial cells involved molecular mechanisms might constitute promising targets for OUD treatment.

PrgmNr 3339 - Identifying and characterizing superhypomethylation in head and neck cancer: drivers and transcriptional effects

View session detail

Author Block: **O. Dancu**[1], X. Chen[2], A. Djedid[3], C. Lu[2], J. Majewski[1]; [1]McGill Univ., Montreal, QC, Canada, [2]Columbia Univ., New York City, NY, [3]St-Justine Hosp., Montreal, QC, Canada
Disclosure Block: **O. Dancu:** None.
Epigenetic dysregulation driving transcriptional changes is a phenomenon occurring in many cancers. The effects of dramatic DNA hypomethylation are poorly understood and remain an active area of research. Of particular interest are the aberrant transcription events that are thought to be associated with a hypomethylated DNA methylation landscape. Indeed, these aberrant transcripts are thought to play a role both in carcinogenesis and in immune response. In the Head & Neck Cancer data from the TCGA, we have discovered a severely hypomethylated sample subset that we have termed as â□□superhypomethylatedâ□□. In this work, through analyzing mutational, transcriptional and DNA methylation data, I identified 1) promising candidate drivers of the superhypomethylation, and 2) the transcriptional aberrations that result from this dramatic loss of DNA methylation.This sample subgroup represents an ideal dataset in which to explore the transcriptional aberrations that occur in severely hypomethylated cancer landscapes.In expanding my investigationâ□□s scope to look for other superhypomethylated cancers, I identified superhypomethylated samples in other cancer types and potential superhypomethylation drivers in those cancers as well. Given that this superhypomethylation phenomenon was more widespread than just occurring in HNSC, this indicates wider clinical relevance than we initially anticipated.

PrgmNr 3340 - Integrative meta-analysis of blood and brain epigenome-wide association studies of Alzheimerâ□□s disease

View session detail

Author Block: **L. Wang**[1], T. Silva[1], J. Young[2], L. Zhang[1], M. A. Schmidt[3], L. Gomez[4], A. Varma[4], E. R. Martin[1]; [1]Univ. of Miami Miller Sch. of Med., Miami, FL, [2]Univ Miami, Miami, FL, [3]Univ Miami Miller Sch Med., Miami, FL, [4]Univ. of Miami, Miami, FL

Disclosure Block: **L. Wang:** None.

Alzheimerâ□□s disease (AD) is the most common cause of dementia and affects about 11% of people 65 years and older in the U.S. With the rising elderly population, AD has become a major public health concern. Currently, there is a severe lack of objective, inexpensive, and minimally invasive biomarkers for AD. Methylated DNA is more stable than mRNA and might provide an excellent source of biomarkers. We performed a meta-analysis of 1620 blood samples from two large blood-based epigenome-wide association studies in AD, generated by the ADNI and AIBL consortiums recently. We found DNA methylation levels at 5 CpGs, mapped to the *SPIDR*, *CDH6* genes, and intergenic regions were significantly associated with AD diagnosis. In addition, to prioritize significant methylation differences associated with AD in both brain and blood, we also performed a cross-tissue meta-analysis by combining these methylation datasets measured in blood with four additional DNA methylation datasets, which included 1030 samples generated from the brain prefrontal cortex, to prioritize 97 CpGs and 10 differentially methylated regions that are significantly associated with both AD neuropathology and AD diagnosis. Integrative analysis with AD GWAS loci, nearby gene expression, TF binding sites, and protein biomarkers suggested these DNA methylation differences play important functional roles in AD. In particular, our analysis revealed that DNA methylation at cg05157625, located in the gene body of the *RIN3* gene, is significantly associated with both AD diagnosis and AD neuropathology as well as expressions of the target gene *RIN3*. *RIN3* is involved in endosomal transport and signaling and has previously been linked to AD in GWA studies. Importantly, DNA methylation at cg05157625 also showed considerable agreement in matched brain and blood samples, indicating it could be a candidate biomarker for AD.

PrgmNr 3341 - Methylation Profiling using Reduced Representation Bisulfite Sequencing for Basal Cell Carcinoma in a Population Exposed to Arsenic

View session detail

Author Block: **G. Da Silva**[1], F. Jasmine[1], M. Argos[2], M. Rakibuz-Zaman[3], T. Islam[3], A. Ahmed[3], M. Islam[3], H. Ahsan[1], M. G. Kibriya[1]; [1]Publ. Hlth.Sci., Univ. of Chicago, Chicago, IL, [2]Univ. of Illinois at Chicago, Chicago, IL, [3]Univ. of Chicago, Bangladesh, Dhaka, Bangladesh
Disclosure Block: **G. Da Silva:** None.
Background: Non-melanoma skin cancers (NMSC) are the most prevalent malignancy in the United States, with an estimated 2 million new diagnoses each year. NMSC includes basal cell carcinoma (BCC) and squamous cell carcinoma. While ultraviolet (UV) radiation exposure and skin sensitivity are known risk factors for NMSC, in Caucasians, arsenic (As) exposure may be a major risk factor in other populations. **Aim**: To identify differential DNA methylation in As-related BCC lesions, we compared DNA from BCC tissue (n=7) and normal skin tissue (n=3). **Material and Methods**: We utilized tissue samples from participants in the **B**angladesh vitamin **E** and **S**elenium **T**rial (BEST). From BEST participants who developed NMSC over the follow-up period of 8 years and had skin biopsy tissue preserved in RNA Later available, we selected the first seven male BCC cases. Healthy skin tissues (surrounding non-cancerous skin lesions) were taken from an independent set of participants (who had arsenical keratosis) for comparison. All of the participants were exposed to As through drinking As-contaminated groundwater and had visible As-related skin lesions at the enrollment of the trial. Over 3.6 million CpG sites were sequenced using Reduced Representation Bisulfite Sequencing Library Kit on the Illumina platform. The Bismark was used for the alignment of the reads and the methylation extraction. SeqMonk was used for the differential DNA methylation analysis. **Results**: From the sequence data, we created a total of 130,040 probes with a window size of 60 CpGs, and the bisulphite pipeline was used. Chi-Square test (q-value Conclusion: We present one of the very first genome-wide DNA bisulfite sequencing studies in BCC in Arsenic exposed males to identify several differentially methylated genes that may be relevant in understanding the pathogenesis of BCC. Further studies are needed to confirm the findings.

PrgmNr 3342 - Methylome-Wide Association Study Identifies Candidate DMRs for Carotid Bifurcation Intima-Media Thickness in Dominican Republic Families

View session detail

Author Block: **N. Dueker**[1], H. Zhao[2], C. Dong[3], D. Cabral[4], R. L. Sacco[4], S. H. Blanton[3], L. Wang[5], T. Rundek[4]; [1]1.John P. Hussman Inst. for Human Genomics, Univ. of Miami, Miami, FL, [2]Yale Univ. Sch. of Publ. Hlth., New Haven, CT, [3]Univ Miami, Miami, FL, [4]Dept. of Neurology, Miller Sch. of Med., Univ. of Miami, Miami, FL, [5]Univ. of Miami, Miami, FL
Disclosure Block: **N. Dueker:** None.
Stroke is a complex disorder influenced by genetic, epigenetic and environmental factors. One strategy to identify stroke risk factors is to study intermediate phenotypes which may be less complex than the clinical outcome. We have previously assembled a large collection of extended families from the Dominican Republic to characterize the genetic determinants of subclinical phenotypes for stroke, including carotid intima-media thickness (IMT). Using this sample, we have shown evidence for linkage and association of genetic variants with carotid bifurcation IMT (BIF), although identified variants only account for a portion of variation in BIF. To identify the â□□missing heritabilityâ□□, we performed a methylome-wide association study to identify methylation sites associated with BIF in 61 extended Dominican families with 799 individuals. CpG methylation in blood-derived DNA was assayed using the Illumina Infinium Human MethylationEPIC BeadChip. Using these data, linear mixed model analyses were performed regressing BIF on beta values for individual CpG sites, adjusting for age, sex, the first ancestry principal component (PC), two methylation PCs and cell composition proportions as fixed effects. Family and batch effects were included as random effects. Region-based analyses were performed using Comb-p to identify differentially methylated regions (DMRs). From this analysis, we identified 15 individual CpGs suggestively associated with BIF (p-5). These included cg04950204, located on chromosome 2 near the hippocalcin like 1 (*HPCAL1*) gene ($p=6.5x10^{-6}$). Notably, methylation within *HPCAL1* was previously implicated in incident myocardial infarction. Region-based analyses identified 24 DMRs significantly associated with BIF, with an average of 4.9 Â± 2.2 CpGs per DMR. Our most strongly associated DMR, chr1: 205819178- 205819610, was located on the peptidase M20 domain containing 1 (*PM20D1*) gene (N CpGs=10; Sidak p-value=$1.8x10^{-11}$). Methylation and genetic variants within *PM20D1* have been previously implicated in obesity and diabetes. Additional DMRs implicated genes involved in processes and phenotypes related to atherosclerosis including adaptive immunity (*PRR5*), chronic kidney disease (*PTPRN2*), Type I and Type II diabetes mellitus (*PTPRN2*, *MAEA*). Together, these results provide suggestive evidence for a role of DNA methylation in subclinical atherosclerosis in Dominicans. Additional analyses exploring the genetic and environmental factors influencing these associations are in progress.

PrgmNr 3344 - PTSD, major depression, and advanced transcriptomic age in neural tissue

View session detail

Author Block: **X. Zhao**[1,2], M. W. Logue[1,2,3,4], Z. E. Neale[1,5], S. E. Hawn[1,2], B. R. Huber[6,7], Traumatic Stress Brain Research Group, M. W. Miller[1,2], E. J. Wolf[1,2]; [1]Natl. Ctr. for PTSD, VA Boston Hlth.care System, Boston, MA, [2]Dept. of Psychiatry, Boston Univ. Sch. of Med., Boston, MA, [3]BioMed. Genetics, Boston Univ. Sch. of Med., Boston, MA, [4]Dept. of Biostatistics, Boston Univ. Sch. of Publ. Hlth., Boston, MA, [5]Dept. of Psychology, Virginia Commonwealth Univ., Richmond, VA, [6]Pathology and Lab. Med., VA Boston Hlth.care System, Boston, MA, [7]Dept. of Pathology, Boston Univ. Sch. of Med., Boston, MA
Disclosure Block: **X. Zhao:** None.
Psychiatric stress has been associated with advanced epigenetic age in DNA methylation in blood and brain tissue, yet no study to date has examined this relationship in the transcriptome. The recent development of a multi-tissue RNA age algorithm (RNAAgeCalc) permits the evaluation of factors that accelerate cellular aging in gene expression. We examined age-adjusted RNA age (AgeAccelRNA) and its association with posttraumatic stress disorder (PTSD), major depressive disorder (MDD), and alcohol use disorder (AUD) using RNA sequence data from three postmortem cortical tissues--dorsolateral prefrontal cortex (dlPFC), ventromedial prefrontal cortex (vmPFC), and motor cortex. Tissue was from the VA National PTSD Brain Bank (n = 55 men, n = 39 women; 74% PTSD and/or MDD; PTSD and MDD cases were grouped due to considerable comorbidity). RNA age was calculated from gene expression values in each brain region and AgeAccelRNA was defined as the residuals from a regression of age at death predicting RNA age. Psychiatric disorders were examined in linear models predicting AgeAccelRNA. In the region(s) that achieved corrected significance, we identified the genes in the RNAAgeCalc that were differentially expressed as a function of the same psychiatric variable(s). We also tested the association between AgeAccelRNA and the neural cell type balance estimated by BrainInABlender. RNA age was highly correlated with age at death across the three regions (rs=.69-.76, ps<.001 ptsd was significantly associated with ageaccelrna in the vmpfc p=".004," padj=".012)" after adjustment for three regions. genes were correction rnaagecalc:>KCNJ16 (log2foldchange [l2fc] = -.84, padj=6.45E-05), *HYAL1* (l2fc = .61, padi=.02), and *CEBPB* (l2fc = .59, padi=.02). Inflammatory and immune-related pathways were enriched among the 43 RNAAgeCalc genes that were at least nominally associated (p

PrgmNr 3345 - Role of differential DNA methylation in Mexican Americans & their risk for Alzheimerâ□□s disease & type 2 diabetes

View session detail

Author Block: **A. Abraham Daniel**[1], T. Silzer[2], J. Sun[1], C. Hall[1], Z. Zhou[3], N. R. Phillips[1], R. C. Barber[1]; [1]UNT Hlth.Sci. Ctr., Fort Worth, TX, [2]Hosp. for Sick Children, Toronto, ON, Canada, [3]Sch. of Publ. Hlth., UNT Hlth.Sci. Ctr., Fort Worth, TX

Disclosure Block: **A. Abraham Daniel:** None.

DNA methylation is an epigenetic mechanism that influences gene expression and has been associated with age related diseases such as metabolic disorders and cognitive decline in several cohorts. Mexican Americans are the largest ethnic minority group in the US and are consequently predicted to have the largest elderly ethnic population within a few decades. Susceptibility to age-related phenotypes such as Alzheimerâ□□s disease (AD), mild cognitive impairment (MCI) and prevalence of type 2 diabetes (T2D), are unique in this cohort. Mexican Americans have an earlier age of onset for cognitive decline (AD/MCI) and have a higher prevalence rate of T2D in comparison to non-Hispanic whites. Mexican Americans also have a metabolic heavy predisposition for AD, compared to non-Hispanic whites who develop inflammation-based AD. The risk for these phenotypes is multifactorial involving epigenetic factors, such as differential DNA methylation, which is the addition of a methyl group to certain cytosine bases of DNA in the genome. Studies have shown the presence of T2D almost doubles the risk of developing AD/MCI, however an epigenetic link between these phenotypes in the Mexican American cohort remains to be established. We aim to elucidate an epigenetic association between cognitive decline and T2D, specific to the Mexican American population, through analysis of methylation profiles from participants selected from the Texas Alzheimer's Research and Care Consortium (TARCC). The TARCC participants consist of Mexican Americans diagnosed with either cognitive impairment alone (either AD/MCI), T2D alone or both cognitive impairment and T2D together and matched for gender and age with a non-Hispanic white counterpart. Peripheral blood samples were drawn from participants and run on the Illumina Infinium MethylationEPIC chip array assessing >850,000 CpG sites to obtain individual methylation profiles. The Chip Analysis Methylation Pipeline (ChAMP) package within R software will be used to assess differential methylation profiles from all 566 TARCC participants. The results obtained will be analyzed using pathway and gene set enrichment analysis tools. Potential identification of differential methylation specific to the Mexican American population could help provide ethnic specific risk information for T2D and cognitive decline together. This could contribute towards developing ethnic-specific biomarkers, treatments or therapies in the near future.

PrgmNr 3346 - Single-cell coexpression analysis of schizophrenia genes refines the dopamine hypothesis and implicates novel treatments

View session detail

Author Block: **K. Pang**[1,2], Z. Liu[1,2]; [1]Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst., Houston, TX
Disclosure Block: **K. Pang:** None.
A significant number of genes have been implicated in schizophrenia, but their contributions to schizophrenia pathology and treatment development are difficult to decipher without understanding their functions in specific brain cell types. Here, we integrated schizophrenia genetics with single-cell RNA sequencing data of dopamine receptors-expressing neurons in the striatum to assess coexpression patterns of different types of schizophrenia genes. We discovered significant negative correlation of schizophrenia genes with the *DRD2* gene which encodes dopamine D2 receptor and the genes whose expression represents the expression state of dopamine D2 receptor-expressing neurons (D2 neurons), indicating high functional convergence of schizophrenia genes in D2 neurons and suggesting convergent dysregulation of D2 neuron function beyond D2 receptor expression in the presence of schizophrenia gene mutations. Further integration with drug-target gene interaction data elucidated mechanisms of action of the approved antipsychotic drugs and prioritized novel high-confidence drug targets and drugs that can potentially modulate the expression state of D2 neurons. The most promising drugs either convergently target D2 receptor, its interacting pathways, and its downstream signaling cascades or hit other individual components crucial for the function of D2 neurons, providing a rich resource to achieve shared and/or personalized therapies tailored to the genetic variants of individual schizophrenia patients. This approach provides a general framework for dissecting the cell-type-specific pathophysiology of complex genetic diseases and identifying genetics-based drugs to target disease mechanisms.

PrgmNr 3348 - Transcriptomic and epigenomic signatures associated with autoimmunity in Parkinson's Disease

View session detail

Author Block: **L. Andriamboavonjy**[1,2], C. Michaud[1,2], A. Lannuzel[3], M. Panisset[1,2], D. Matheoud[1,2], M. TÃ©treault[1,2]; [1]Univ. of Montreal, Montreal, QC, Canada, [2]CRCHUM, MontrÃ©al, QC, Canada, [3]Inst. du cerveau et de la moelle Ã©piniÃ¨re, Paris, France
Disclosure Block: **L. Andriamboavonjy:** None.
I- Objectives: 1- To characterize the transciptomic and/or epigenomic modulation of autoimmunity associated genes in parkinson's disease 2- To identify the potential omics signature involved in parkinsonism clinical expression (Parkinson's Disease vs. other parkinsonism)
II- Background: The correlation between autoimmunity and Parkinson's disease pathogenesis tend to be confirmed. In the familial form of PD, genetics analysis has confirmed the association between PD linked genes and autoimmune diseases. Furthermore, PINK1 and PARKIN proteins involved in PD are essential in immune response via the mitochondrial antigen presentation pathway. The autoimmune basis in PD has led us to the hypothesis of a peri-genetic modification in autoimmunity associated genes for sporadic PD. Besides, transcriptomic and epigenomic alteration can occur after a viral infection or downstream to the activation of immune cells. In Guadeloupe, where flaviviruses infections are endemic, the prevalence of atypical parkinsonism is much higher compared to North America and Europe. We suspect a link between the flaviviruses infection and atypical parkinsonism due to a peri-genetic alteration.
III- Methods: Six specific groups will constitute our research population:- From Montreal: PD patients, atypical parkinsonism patients, control patients
- From Guadeloupe: PD patients, atypical parkinsonism patients, control patients
1- Establishing a transcriptomic signature: Peripheral blood mononuclear cells' (PBMC) RNA will be extracted and subject to a bulk RNA-seq. A bioinformatic pipeline will align the transcripts and identify: variants, alternative splicing, differential expression, specific isoforms and differential polyadenylation.
2- Establishing an epigenomic signature: PBMC's DNA will be extracted and subjected to a methyl-ATAC-sequencing, which will allow to determine the opened chromatin regions and the methylation profile.
IV- Results: Research on process.
V- Conclusion: Presently, our understanding of the neuropathogenesis of PD is not well-rounded. The path of autoimmunity has to be deepened. Exploring the genetic and peri-genetic differences between PD and other parkinsonism can lead us to a more accurate diagnosis and a better care for patients with this neurodegenerative disease.

PrgmNr 3349 - A cryptic splice acceptor as a cause of X-linked heterotaxy

View session detail

Author Block: **J. Wells**[1], M. B. Padua[2], S. M. Ware[1,2]; [1]Dept. of Med. and Molecular Genetics, Indiana Univ. Sch. of Med., Indianapolis, IN, [2]Dept. of Pediatrics, Indiana Univ. Sch. of Med., Indianapolis, IN
Disclosure Block: **J. Wells:** None.
Heterotaxy syndrome is typically characterized by severe congenital heart defects (CHDs) and abnormal arrangement of thoracic and abdominal organs due to perturbed left-right (LR) asymmetry. The genes currently known to cause heterotaxy explain only 15-20% of cases and are largely biologically related to the structure or function of the embryonic node, a ciliated LR organizer. Mutations in Zinc finger of the Cerebellum 3 (*ZIC3*) are the only known cause of X-linked heterotaxy, accounting for 75% of X-linked cases. We identified an X-linked heterotaxy pedigree with four affected males and performed exome sequencing. Although we did not identify a coding variant in *ZIC3*, a rare predicted damaging hemizygous variant in G-protein coupled receptor 101 (*GPR101* c.1225G>A;p.V409M), a gene tightly linked to *ZIC3*, segregated with disease in the pedigree. *GPR101* is a compelling candidate based on the known cilia function of its closely related family member *GPR161*. Morpholino-induced knockdown of *gpr101* in Xenopus laevis resulted in a heterotaxy phenotype. However, our *Gpr101* global knock-in mouse line with an inserted LacZ reporter (*Gpr101*$^{tm1b(KOMP)Mbp}$) did not display a heterotaxy phenotype and expression analysis showed that *Gpr101* is not expressed at the embryonic node. In the same X-linked heterotaxy pedigree, we conducted whole genome sequencing and identified a novel, intronic variant in *ZIC3* also segregating with disease phenotype. This new *ZIC3* c.1224+3286A>G variant was predicted to result in a cryptic splice acceptor that would lead to novel exon inclusion after exon 2. Structurally, this novel isoform is predicted to share 408 amino acids with known *ZIC3* isoforms followed by cryptic exon inclusion encoding 80 novel amino acids and several predicted polyadenylation sites. This novel isoform is predicted to lack a critical portion of the fifth zinc finger DNA binding domain previously demonstrated to be required for DNA binding and nuclear localization. A minigene splicing assay was performed in HEK293 cells to determine if cryptic exon inclusion occurs. This assay confirmed that the *ZIC3* c.1224+3286A>G variant results in altered splicing between exon 2 and the predicted cryptic exon. Experiments are in progress to further functionally validate this novel isoform. These results demonstrate the second reported instance of a noncoding *ZIC3* variant associated with heterotaxy and the first due to cryptic exon inclusion.

PrgmNr 3350 - Adipose eQTL meta-analysis of 2,256 samples identifies 517 conditionally distinct signals colocalized with 426 cardiometabolic trait GWAS loci

View session detail

Author Block: **S. M. Brotman**[1], L. Guan[2], J. S. El-Sayed Moustafa[3], A. U. Jackson[4], K. A. Broadaway[1], C. K. Raulerson[1], M. R. Erdos[5], N. Narisu[5], P. Pajukanta[6], M. Boehnke[4], M. Laakso[7], F. S. Collins[5], K. L. Mohlke[1], L. J. Scott[4], K. S. Small[3]; [1]Dept. of Genetics, Univ. of North Carolina, Chapel Hill, NC, [2]Dept. of Computational Med. & Bioinformatics, Univ. of Michigan, Ann Arbor, MI, [3]Dept. of Twin Res. and Genetic Epidemiology, King's Coll. London, London, United Kingdom, [4]Dept. of Biostatistics and Ctr. for Statistical Genetics, Sch. of Publ. Hlth., Ann Arbor, MI, [5]Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [6]Dept. of Human Genetics and Inst. for Precision Hlth., David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [7]Inst. of Clinical Med., Kuopio Univ. Hosp., Univ. of Eastern Finland, Kuopio, Finland
Disclosure Block: **S.M. Brotman:** None.
At many GWAS loci, the genes underlying disease risk remain unknown. Candidate genes can be detected through colocalization of GWAS loci with expression quantitative trait loci (eQTL) in trait-relevant tissues, and colocalized loci may be more readily identified by analyzing conditionally distinct GWAS and eQTL signals. We performed an RNA-seq-based local (Â± 1 Mb) eQTL meta-analysis of 2,256 subcutaneous adipose tissue samples from 5 studies, and performed conditional analysis to detect distinct signals using APEX. Of 28,346 genes tested in â¥2 studies, 19,286 (68.0%) had â¥1 significant eQTL, which is more than the 13,115-17,582 genes with eQTL in the individual studies. Of these 19,286 genes, 10,244 (53.1%) genes had â¥2 eQTL signals and 4,562 (23.7%) genes had â¥3 eQTL signals. Among genes with â¥2 eQTL signals, we find a median distance of 28 kb from the lead variant to the transcription start site (TSS) for primary signals, 41 kb for secondary signals, and 59 kb for 3rd to 10th signals. As the number of signals increased, the median eQTL effect size generally decreased: 0.26 for primary signals, 0.17 for secondary signals, and 0.17 for 3rd to 10th signals. Compared to the 10-fold larger eQTLGen blood eQTL meta-analysis (n > 30,000 samples), more than half of adipose eQTL signals (20,372; 54.0%) were not significant (FDR coloc2 (PP4 > 0.8) to perform colocalization of adipose eQTL with GWAS loci from six cardiometabolic traits. In total, 426 GWAS loci (132 waist-to-hip ratio adjusted for BMI, 115 body mass index, 110 waist-to-hip ratio, 61 type 2 diabetes, 4 coronary artery disease, 4 fasting glucose) were colocalized with eQTL signals from 517 genes. Secondary eQTL signals have not been analyzed in most prior studies, but here we observed that more secondary eQTL signals colocalized with GWAS loci. Among the genes with â¥2 eQTL, 171 primary eQTL were colocalized with 159 GWAS loci, while 221 eQTL with â¥2 signals were colocalized with 193 GWAS loci. One example of an eQTL detected in the meta-analysis but not the individual studies is for *EXOC3L1* (eQTL *p*-value = 8.1 x 10^{-8}), which is colocalized with a waist-to-hip ratio adjusted for BMI GWAS signal. *EXOC3L1* encodes a protein of the exocyst complex that functions as a tether of secretory vesicles to the plasma membrane to facilitate molecular trafficking. Using our five-study adipose eQTL meta-analysis, we find 53.1% of genes have conditionally distinct eQTL signals that would have been missed if we only detected primary signals, and we identify biologically plausible genes that may influence variation of cardiometabolic traits.

PrgmNr 3351 - Characterization of neonatal epigenetic markers associated with elevated stress from the COVID-19 pandemic

View session detail

Author Block: **K. Kocher**[1,2], S. Bhattacharya[1], J. Quistorff[3], C. Lopez[3], M. Almalvez[1], N. Niforatos-Andescavage[3], E. J. Vilain[1,2], C. Limperopoulos[3], E. Delot[1,2]; [1]Ctr. for Genetic Med. Res., Children's Natl. Med. Ctr., Washington, DC, [2]Dept. of Genomics & Precision Med., George Washington Univ., Washington, DC, [3]Developing Brain Inst., Children's Natl. Hosp., Washington, DC
Disclosure Block: **K. Kocher:** None.
With the sudden outbreak of COVID-19 in early 2020 came a prolonged period of quarantine and shutdown of society. Job insecurity, economic instability, isolation, and fear of exposure to COVID-19 led to widespread psychological distress for many. During gestation, stressors to the fetus, including maternal stress, anxiety, and depression, can lead to impaired fetal brain development, which may result in long-term reduced quality of life and cognitive impairment. However, the mechanisms by which these prenatal stressors disrupt fetal brain development remain unclear, and outcomes vary even for similar stressors, suggesting the existence of factors providing resilience or susceptibility to stress. We have previously identified DNA methylation patterns in newborn issued from stressed pregnancies of other causes (e.g. heart or placental defect, premature birth). In this study we aim to determine whether there are unique epigenetic signatures in newborns associated with stress during pregnancy due to the COVID-19 pandemic that may accompany modified fetal brain development. We investigated DNA methylation differences in newborns who experienced otherwise healthy pregnancies that occurred during the COVID-19 pandemic (RESCUE cohort). Biospecimens were collected at birth as part of a comprehensive, longitudinal study. Maternal mental health metrics as well as fetal brain MRI imaging were recorded throughout pregnancy and at birth. We designed and validated a bioinformatic pipeline to filter and normalize output data from Illumina Infinium MethylationEPIC BeadChip for this application. Unsupervised clustering showed appropriate clustering by tissue type, no gender bias, and gross appropriate clustering of cohort samples. Exceptions to this clustering are being further studied to uncover phenotypic associations using clinical, demographic, and MRI data. Widespread differential methylation was found between RESCUE (n=33) newborns and age- and sex-matched healthy controls (CTL, n=13), recruited before the start of the pandemic. In contrast, there were no apparent epigenetic differences associated with COVID-19 infection during pregnancy. While we found specific DNA methylation signatures that are unique to this cohort, we also identified signatures that overlap between other stressful pregnancy cohorts, suggesting that there may be common epigenetic markers to all types of prenatal stress.

PrgmNr 3352 - Defining the effects of noncoding genetic variation on human regulatory element activity

View session detail

Author Block: **K. Strouse**, G. D. Johnson, K. Siklenka, A. Barrera, W. Majoros, A. S. Allen, T. E. Reddy; Duke Univ., Durham, NC
Disclosure Block: **K. Strouse:** None.
Millions of noncoding genetic variants are associated with common human traits and disease. As a step towards understanding how noncoding variants contribute to disease, we are generating a population-scale atlas of the effects of 40 million genetic variants on gene regulatory element activity. To do so, we are assaying regulatory activity genome-wide across 300 individuals from diverse populations using the high-throughput reporter assay STARR-seq. We have developed a pooling approach to whole-genome STARR-seq that allows for comprehensively assaying regulatory variation across five individuals at once; and a Bayesian model of allele-specific STARR-seq signal to estimate allele-specific regulatory activity from that data. To maximize the number of genetic variants we assay, we have prioritized the genomes of individuals with African ancestry, but still include representation from diverse ancestries. In total, we will assay all common noncoding variants and over 20 million rare variants (MAF

PrgmNr 3353 - DNA-binding consequences of human homeodomain missense variants

View session detail

Author Block: **K. Kock**[1], P. K. Kimes[2], S. Inukai[1], S. K. Phanor[1], S. S. Gisselbrecht[1], C. H. Lipper[3], G. Ramakrishnan[1], J. T. Anderson[1], J. V. Kurland[1], J. M. Rogers[1], S. C. Blacklow[3], R. A. Irizarry[2], M. L. Bulyk[1]; [1]Brigham and Women's Hosp. and Harvard Med. Sch., Boston, MA, [2]Dana-Farber Cancer Inst. and Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [3]Dana-Farber Cancer Inst. and Harvard Med. Sch., Boston, MA

Disclosure Block: **K. Kock:** None.

Homeodomains are transcription factors with important roles in development, patterning, and cellular differentiation. Numerous homeodomain coding mutations, particularly within the DNA-binding domain (DBD), have been associated with an array of human Mendelian diseases. While a select number of these mutations have been studied in detail, the broader biochemical consequences of missense variation across the span of the homeodomain DBD remain understudied. Since every single position within the homeodomain DBD has been reported in ClinVar to have pathogenic variants and variants of uncertain significance (VUS), a more comprehensive understanding of the effects of these variants is important for variant interpretation and insights into disease mechanisms.

We sought to elucidate the consequences of such missense variation on DNA binding by studying disease-associated and naturally-occurring variants, alongside variants designed to test for tolerance of mutations at particular residues. We profiled >90 DBD missense variants across 30 human homeodomains using protein-binding microarrays to evaluate DNA binding to all possible 8 base pair sites (8-mer). We utilised a parametric model framework to make inferences about differential binding affinity and altered binding specificity by homeodomain missense variants to sets of 8-mers.

More than 75% of disease-associated variants we examined resulted in diminished binding affinity, particularly to 8-mers bound by the wild-type homeodomain with high affinity. Several variants showed stark changes in DNA-binding specificity, with recognition of novel sets of 8-mers. More surprisingly, we identified variants that altered binding affinity specifically for sets of 8-mers that were recognised by the wild-type homeodomain at moderate affinity, which we interpreted as a subtler change in DNA-binding specificity. We identified 9 novel DNA-binding specificity-determining positions across our dataset, with several distal from homeodomain-DNA interfaces. We performed structural modelling to suggest mechanisms for alterations in DNA-binding specificity, such as changes in intra-homeodomain side-chain interactions, or introduction of contacts with DNA target sites.

Our results suggest pathogenic mechanisms for several VUS and disease-associated variants and genes, and contribute towards clarification of VUS; ~5% of homeodomain DBD VUS in ClinVar affect novel specificity-determining positions identified in our study. We envision our findings contributing towards a deeper understanding of homeodomain-DNA recognition rules, and improved variant interpretation and target gene prediction.

PrgmNr 3354 - Genome-wide association studies identify novel genetic loci for epigenetic age acceleration among survivors of childhood cancer

View session detail

Author Block: **Q. Dong**[1], N. Song[2,3], N. Qin[4,5], Z. Li[1], X. Sun[1], J. Easton[2], H. Mulder[1], G. Neale[1], E. Walker[1], X. Ma[1], I-C. Huang[1], J. Zhang[6], K. K. Ness[1], M. M. Hudson[1], L. L. Robison[1], Z. Wang[6]; [1]St. Jude Children's Res. Hosp., MEMPHIS, TN, [2]St. Jude Children's Res. Hosp., Memphis, TN, [3]Chungbuk Natl. Univ., Cheongju, Korea, Republic of, [4]Nanjing Med. Univ., Nanjing, China, [5]St. Jude Childrenâ□□s Res. Hosp., Memphis, TN, [6]St Jude Children's Res. Hosp., Memphis, TN

Disclosure Block: **Q. Dong:** None.

Background: We have shown that epigenetic age acceleration (EAA), based on Levineâ□□s clock (i.e., PhenoAge), is significantly higher in survivors of childhood cancer and is associated with specific treatment exposures, unfavorable health behaviors, and presence of specific chronic health conditions. To better understand inter-individual variability, we aimed to investigate the genetic basis of EAA among survivors of childhood cancer.

Methods: Genome-wide association studies of EAA based on four different epigenetic clocks (Hannum, Horvath, PhenoAge, and GrimAge) were performed with DNA methylation EPIC array and whole-genome sequencing data generated from blood-derived DNA from 2138 survivors (discovery), an additional set of 504 survivors (replication), and 282 community controls for comparison (exploration). Both sets of survivors as well as community controls are of European ancestry from the St. Jude Lifetime Cohort Study. Linear regression models were fit for each epigenetic age against the additive dose of each genetic variant across the genome, adjusting for age at DNA sampling, sex, and cancer treatment exposures. Fixed-effects meta-analysis was used to combine summary statistics for discovery and replication sets. LD score regression was used to estimate SNP-based heritability.

Results: For EAA-Horvath, a genome-wide significant association peak was mapped to *SELP* gene with the strongest SNP rs732314 (combined: Beta=-0.56, P=3.91Ã□10^{-11}, I^2=50.48%; discovery: Beta=-0.50, P=1.55Ã□10^{-7}; replication: Beta=-0.81, P=2.69Ã□10^{-5}). rs732314 is a meQTL for cg01459453 (a CpG in Horvathâ□□s clock). For EAA-Hannum, an association peak was mapped to *HLA* locus with the strongest SNP rs28366133 (combined: Beta=-0.78, P=3.74Ã□10^{-11}, I^2=0; discovery: Beta=-0.76, P=5.30Ã□10^{-8}; replication: Beta=-0.84, P=2.02Ã□10^{-4}). There was no genome-wide significant SNP for EAA-PhenoAge or EAA-GrimAge. Interestingly, among community controls, rs732314 was significantly associated with EAA-Horvath (Beta=-1.09, P=5.43Ã□10^{-5}) but rs28366133 was not associated with EAA-Hannum (Beta=-0.21, P=0.48). The estimated heritability was 0.33 (SE=0.20) for EAA-Horvath and 0.17 (SE=0.23) for EAA-Hannum. In contrast, the estimated heritability was close to zero for EAA-PhenoAge and EAA-GrimAge.

Conclusions: We identified novel genetic variants in *SELP* and *HLA* gene loci associated with EAA-Horvath and EAA-Hannum, respectively, among survivors of childhood cancer. Heritability estimates showed measurable genetic components of EAA based on the first generation of epigenetic clocks (Horvath and Hannum) but not the second generation (PhenoAge and GrimAge).

PrgmNr 3355 - Genome-wide detection of enhancer-hijacking and reconstructing complex SVs in cancer genome

View session detail

Author Block: **F. Yue**, X. Wang, J. Xu, B. Zhang, Y. Hou, F. Song, H. Lyu; Northwestern Univ., Chicago, IL
Disclosure Block: **F. Yue:** Major Stockholder/Ownership Interest; Sariant therapeutics. Receipt of Intellectual Property Rights/Patent Holder; Feng Yue and Xiaotao Wang.
Recent efforts have shown that structural variations (SVs) can disrupt three-dimensional genome organization and induce enhancer hijacking, yet no computational tools exist to identify such events from chromatin interaction data. Here, we develop NeoLoopFinder, a computational framework to identify the chromatin interactions induced by SVs, including interchromosomal translocations, large deletions and inversions. Our framework can automatically resolve complex SVs, reconstruct local Hi-C maps surrounding the breakpoints, normalize copy number variation and allele effects and predict chromatin loops induced by SVs. We applied NeoLoopFinder in Hi-C data from 50 cancer cell lines and primary tumors and identified tens of recurrent genes associated with enhancer hijacking. To experimentally validate NeoLoopFinder, we deleted the hijacked enhancers in prostate adenocarcinoma cells using CRISPR-Cas9, which significantly reduced expression of the target oncogene. In summary, NeoLoopFinder enables identification of critical oncogenic regulatory elements that can potentially reveal therapeutic targets.

PrgmNr 3356 - Automated localization and quantification of RNA transcripts from RNA-fluorescent in situ hybridization imaging data

View session detail

Author Block: **B. Hospelhorn**[1], B. K. Kesler[2], G. Neuert[2,3,4]; [1]Vanderbilt Genetics Inst., Vanderbilt Univ., Nashville, TN, [2]Dept. of Molecular Physiology and Biophysics, Vanderbilt Univ., Nashville, TN, [3]Dept. of Pharmacology, Vanderbilt Univ., Nashville, TN, [4]Dept. of BioMed. Engineering, Vanderbilt Univ., Nashville, TN

Disclosure Block: **B. Hospelhorn:** None.

The regulation of gene expression is a major factor in cells' ability to respond to environmental stimuli and differentiate during the development of multi-cellular organisms. However, a transcriptional program is not only defined by which genes are upregulated and downregulated. Understanding the finer kinetics of transcriptional tuning is important for fully understanding transcription as a process and how its regulation impacts cellular phenotype. Because variability increases with length of time and the number of cells measured in a population, high resolution spatiotemporal localization of individual RNA molecules in single cells is important for studying these dynamics at high precision. RNA fluorescent in situ hybridization (RNA-FISH) is a frequently used technique for visualizing RNA in fixed cells using fluorescent probes. Automated processing of the resulting images is essential for large datasets that may include many different transcripts and time points. Here we demonstrate that our MATLAB based RNA-FISH image processing pipeline is a useful tool for automatically detecting the 3D locations of cell boundaries and RNA transcripts at single molecule resolution in an RNA-FISH image stack. In particular, this tool is effective for facilitating quantitative analyses of FISH data such as determining the colocalization of multiple transcripts or the relative amount of RNA in various subcellular compartments. We aim to develop a tool capable of performing analysis for a variety of imaging-based approaches to studying RNA dynamics. Our hope is to use this tool to advance understanding of eukaryotic transcription regulation mechanisms by studying the quantities and behaviors of transcripts and regulatory RNA at discrete time points.

PrgmNr 3357 - Deep Learning Study to Predict Ribosome Binding from RNA-seq Data

View session detail

Author Block: **H. Kumar**[1], T. Yang[1], P. Kim[2]; [1]Sch. of BioMed. Informatics, The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX
Disclosure Block: **H. Kumar:** None.
It is critical to have a ribosome footprint for identifying the actively translated mRNAs and peptide folding and measuring the protein synthesis. Ribo-seq provides the global snapshot of the translation efficiency. However, there are not many publicly available Ribo-seq data compared to the huge amount of RNA-seq data. Mathematical modeling of translation process uses such information as production of mRNA, initiation of translation, assembly of ribosome, elongation, termination, degradation of mRNA and proteins. Based on this context, there were several trials to model the translation process from the molecular unit in the view of biochemistry for the prokaryotes. To date, there were no studies to infer the translation status from the RNA-seq data in humans, which is resulted in the relatively low number of studies of translatome compared to the transcriptome. To fill this gap, we will build a deep learning model to infer the active ORFs from RNA-seq data, named virtual Ribo-seq, by utilization of diverse translation-related information from the transcript level such as expression, splicing, and RNA degradation events. We trained our model using publicly available 11 human cell lines that have matched data of RNA-seq and Ribo-seq, incorporating three RNA process information (expression, splicing, and degradation of RNA) and other basic information (conservation and number of evidences of Ribo-seq peak). Virtual Ribo-seq is a convolutional neural network model. We tested our model on one of the most studied human cancer cell-line, K562. This study is mainly aimed to better understand the translation process through inferencing the translation state of individual genes from the active RNA processing status in the new cell.

PrgmNr 3358 - Interpreting Epigenetic Aging with Deep Neural Networks Using High Dimensional DNA Methylation Data

View session detail

Author Block: **H. Guan**; Basis San Antonio Shavano, San Antonio, TX
Disclosure Block: **H. Guan:** None.
Aging is a normal process linked to specific patterns and changes in the epigenome, e.g., DNA methylation. Multiple epigenetic aging clocks have been developed, but little research have focused on interpretation of these aging equations. Deep neural networks (DNN) are powerful prediction models which can model higher-order of interaction between biomarkers. We have previously developed a DNN-based prediction model for aging prediction. In this study, we use three model-agnostic interpretability methods to interpret how trained DNN predict human biological age by high-dimensional DNA methylation data. Specifically, we compare the Local Interpretable Model-agnostic Explanations, Shapley Values, and MMD-Critic by using the maximum mean discrepancy and large-scale submodular optimization. We discover that, for a large number of candidate features, such as genome-wide DNA methylation data, a key factor in improving prediction accuracy is to appropriately weight features that are highly correlated with the outcome of interest. We also show that explanations can mitigate the impact of misclassified features from the perspective of the end-user. We conclude that explanations are valuable for improving our understanding of epigenetics and what it means for aging.

PrgmNr 3359 - Investigating the utility of sex-informed PrediXcan models in whole blood

View session detail

Author Block: **V. A. Janve**, L. Dumitrescu, K. A. Gifford, K. R. Pechman, A. L. Jefferson, T. J. Hohman; Vanderbilt Univ. Med. Ctr., Nashville, TN
Disclosure Block: **V.A. Janve:** None.
Introduction: Based on Oliva et al. (Oliva et al. Science 2020), there are known sex differences in tissue-specific gene expression and in the genetic architecture of gene expression, but such differences have not been incorporated into predicted gene expression models to date. Gene based methods such as PrediXcan use expression quantitative trait loci (eQTLs) to build gene expression models. Compared to traditional GWAS methods these have higher power, reduced multiple testing burden and potential to identify causal pathways. Sex and tissue specific predicted expression can be imputed for investigating the genetic architecture of gene expression, when only genetic data is available. However, careful evaluation is needed before such methods are applied with confidence. **Methods:** We built sex-stratified models using whole blood RNA sequencing in GTEx(v8) and evaluated their performance in an independent dataset. Specifically, prediXcan models were built following the method described in Gamazon et al. (Gamazon et al. Nature Genetics 2015), but we included joint and sex-specific models. Validation was evaluated leveraging RNA sequencing data from Non-hispanic white, cognitively normal partcipants in the Vanderbilt Memory & Aging Project (Jefferson et al. J Neurology 2017, n=170; 99 males, age=72.1Â±6.44; 71 females, age=73Â±8.17). The associations were evaluated with observed expression in all the predicted genes (n=16,592) and sex-biased genes identified in Oliva et al. (n=353 of 500, 189 females and 164 males specific, respectively) in a joint and sex-stratified manner. **Results:** Sex-stratified models perform well, comparable to joint models with more heritable genes showing stronger observed R^2 in validation analyses. Several of PrediXcan gene models are observed in sex-stratified models that are not observed in joint models, however, these sex-specific predicted models inclined not reproduce in a robust manner (only 63/189 reproduced in females and 63/164 in males). When comparing joint models restrictively applied in one sex to sex-stratified models applied in one sex, the joint models tended to perform better, even when restricting to genes that show pronounced sex differences across sex. **Conclusion:** While we do see evidence of sex-specificity in the genetic architecture of gene expression, we do not provide evidence supporting the application of sex-stratified PrediXcan builds across the autosomes in this small study. Future work will integrate the X-chromosome, test models in larger brain eQTL databases that may have increased power for sex-stratified PrediXcan model evaluation.

PrgmNr 3360 - A site-specific mutation rate model uncovers signatures of localized hypermutability

View session detail

Author Block: **V. Seplyarskiy**[1], E. Koch[1], J. S. Lichtman[2], H. H. Luan[2], D. Lee[1], S. Sunyaev[3]; [1]Harvard Med. Sch. Div. of Genetics, Boston, MA, [2]NGM biopharmaceuticals, South San Francisco, CA, [3]Harvard Med. Sch. and Brigham & Women's Hosp., Boston, MA
Disclosure Block: **V. Seplyarskiy:** Consultant/Consulting Fees/Other Remuneration; NGM biopharmaceuticals.
Mutation rate is inhomogeneous along the human genome. The precise knowledge of mutation rate in each position of the genome would highlight relevant mutagenic forces, it would generate a proper baseline for estimation of selective constraint and will equip statistical methods for identifying genes causing human Mendelian diseases. Capitalizing on large-scale population sequencing datasets, we developed a statistical model of mutation rate at the base pair resolution. Our model was trained on the filtered and recurrence corrected rare SNV data from the GNOMAD v.3.1 resource. Validation of the model on *de novo* mutation data obtained in parent-child trio sequencing studies demonstrated high accuracy exceeding that of the existing models. The mutation rate model relies on known genomic factors influencing mutagenesis. It includes extended nucleotide sequence context, enumerating mutagenic properties of each pentanucleotide and using a sparse regression model to incorporate the influence of eight additional nucleotide positions. The model also takes into account DNA methylation, directions of transcription and replication, and mutation rate variation at the 50kb scale for all tri-nucleotide mutation types. We found that mutation rate varies within each pentamer at least 2.5-fold, and up to 100-fold. This implies, that incorporation of the effect of immediately adjacent nucleotides is insufficient to estimate position-specific mutation rate. We detected a number of 100-nucleotide long elements with greatly elevated mutation rate compared to the model. These elements include all tRNA, small nuclear RNA and a subset of ribosomal RNA genes. At the same time, pseudogenes from the same families show no signs of hypermutability. We validated hypermutability of these RNA genes using *de novo* mutations and confirmed the 32-fold increase (18-48, 95% confidence interval) of mutation rate for small nuclear RNAs. A tempting interpretation is that the hypermutability is associated with transcription by polymerase III. For most functional annotations (transcription factor binding sites and ENCODE regulatory elements) the observed mutation rate is in line with our predictions. However, mutation rate in CTCF binding sites in promoters is deviated from our estimates and displays an excess of T>G and T>A mutations. Mutational footprint of CTCF binding in promoters is greatly distinct from both CTCF binding sites outside of promotors and from non-CTCF bound promoters. This enables accurate prediction of CTCF bound promotors solely from mutational data AUC (0.87).

PrgmNr 3361 - Assessment of the rates of pathogenic variants across the All of Us Research Program cohort

View session detail

Author Block: **E. Venner**[1], K. Walker[1], D. Kalra[1], B. Lee[1], P. E. Empey[2], J. H. Karnes[3], J. D. Smith[4], S. R. McGee[5], A. Lewis[6], S. Kalla[1], K. Patterson[4], A. Radhakrishnan[7], A. Haddad[2], M. M. Wheeler[4], Q. Wang[8], D. A. Nickerson[9], D. M. Toledo[10], A. Musick[11], R. A. Gibbs[12]; [1]HGSC - Baylor Coll. of Med., Houston, TX, [2]Univ. of Pittsburgh, Pittsburgh, PA, [3]Univ Arizona, Tucson, AZ, [4]Univ of Washington, Seattle, WA, [5]Univ. of Washington Genome Sci., Seattle, WA, [6]Columbia Univ. Med. Ctr., New York, NY, [7]Univ. of Washington, Seattle, TX, [8]Baylor Coll. of Med., Houston, TX, [9]Univ Washington Sch Med, Seattle, WA, [10]Dartmouth-Hitchcock Med. Ctr., Lebanon, NH, [11]NIH, Bethesda, MD, [12]Baylor Coll. Med., Houston, TX
Disclosure Block: **E. Venner:** Major Stockholder/Ownership Interest; Codified Genomics.
The All of Us Research Program (AoURP) has a more diverse cohort of participants than within previously studied large cohorts. It is currently unknown, however, whether the rate of positive health-related genetic findings in the AoURP cohort will differ from those in the other healthy populations that have been ascertained. Identification of such differences will be a powerful reinforcement of the importance of the AoURPâ□□s mode of recruitment and engagement.
Using the AoU Research Hub, we have begun a â□□demonstration projectâ□□ that will assess health-related genetic findings within AoURP whole genome sequence (WGS) data, and provide feedback on the Research Hub itself. After automated filtering, we have begun manual assessment of putative pathogenic variants on aggregated data from 9,151 AoURP participants. Preliminary results show a 0.5% positive rate for cancer or tumor diseases (not counting recessive conditions MYH-associated polyposis and Wilson disease, which have not been fully assessed during the aggregate review phase), 0.7% for cardiac disease-related findings, and 0.3% for hypercholesterolemia. As these data mature they can be compared to other cohorts and broken down into subgroups by ancestry backgrounds. As a preliminary comparison, in the eMERGE III program, in patients not ascertained for a specific indication, 2% of patients had a finding related to cancer risk, 1.7% had a finding related to cardiac disease, and 1.1% had a finding related to familial hypercholesterolemia. These rates are generally higher than within the AoURP data, but this may reflect that although the eMERGE III participants were not ascertained with a specific indication, they were enrolled as part of clinical visits. These ratios will also change as we continue to review variants of unknown significance.
More studies of specific loci and comparison with other cohorts will determine whether there are exceptional features within the AoURP cohort. This will test the hypothesis that the increased diversity of the AoURP cohort will lead to a higher positive rate of health related results, across the program, than one would expect from a less-diverse population of healthy individuals.

PrgmNr 3362 - Cataloging human *PRDM9* allelic variation using long-read sequencing reveals *PRDM9* population-specificity and two distinct groupings of related alleles

View session detail

Author Block: **B. Alleva**[1], K. M. Brick[2], F. Pratto[3], M. Huang[3], R. D. Camerini-Otero[4]; [1]NIDDK, Bethesda, MD, [2]NIDDK, NIH, Bethesda, MD, [3]NIH, Bethesda, MD, [4]Natl. Inst.s Hlth, Bethesda, MD
Disclosure Block: **B. Alleva:** None.
In humans, localization of DNA double strand breaks (DSBs) during meiosis is dependent upon the DNA-binding protein PRDM9. DSB targeting specificity is encoded by a hypervariable array of C2H2 zinc fingers, resulting in potentially countless alleles with different DNA recognition sites. Since *PRDM9* alleles have gross differences in recombination hotspot localization, systematic assessment of *PRDM9* diversity is important for understanding the complexity of population genetics and inheritance linkage patterns. However, due to the repetitive nature of *PRDM9*, large-scale genotyping is difficult. In addition, previous studies of *PRDM9* diversity disproportionately surveyed individuals of European descent and therefore a comprehensive survey of *PRDM9* across human populations has never been performed. To address these issues, we developed a high throughput long-read sequencing strategy to genotype *PRDM9* from 722 individuals across 7 populations composed of individuals from Asia, Africa, South America, and Europe. We detected 70 alleles including 34 novel alleles and 11 alleles previously identified in sperm suggestive of greater undiscovered *PRDM9* diversity. To assess potential relatedness between human *PRDM9* alleles, we developed an algorithm to infer sequence similarity between different alleles based on template switching as the mechanism for new *PRDM9* allele formation. Our relatedness analysis revealed that these alleles group into two major clusters around the most prevalent alleles, *PRDM9-A* and *PRDM9-C*. Importantly, by mapping DSBs in human testis we also found that individuals harboring small variations in PRDM9 can substantially alter the meiotic recombination landscape. In summary, our data expands our knowledge of *PRDM9* diversity in humans and demonstrates that minor *PRDM9* variants may play an under-appreciated role in shaping patterns of human recombination.

PrgmNr 3363 - Characterization of the Evolutionary Forces Acting on Genomic Regions Associated with over 600 Human Complex Traits

View session detail

Author Block: **A. LaBella**[1], A. Abraham[1], A. Rokas[1], J. A. Capra[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Univ. of California San Francisco, San Francisco, CA
Disclosure Block: **A. LaBella:** None.
Natural selection has shaped the genetic architecture of many human traits. While the action of specific evolutionary forces on some traits has been identified, the role of different modes of selection across the human phenome remains poorly understood. Genome-wide association studies (GWAS) have greatly advanced our understanding of the genetic architecture of diverse complex traits. Meanwhile, genome-wide scans for diverse signatures of selection (GWSS) have also characterized the imprint of evolutionary forces, including different modes of selection, across the human genome. To address this gap, we developed an approach called MOSAIc that integrates these two approaches (GWAS + GWSS), enabling the investigation of signatures of diverse evolutionary forces on trait-associated genomic regions. Application of our MOSAIc approach to loci identified through a GWAS on spontaneous preterm birth revealed a mosaic of evolutionary forces acting on genomic regions associated with spontaneous preterm birth (LaBella, Abraham, *et al.* 2020, *Nature Comm.* 11: 3731). To gain broader insight into the diversity (or lack thereof) of evolutionary forces shaping complex human traits, we applied our MOSAIc approach to >600 GWAS on a multitude of complex human traits, including diseases across multiple domains, and 11 evolutionary metrics based on sequence constraint, populations differentiation, and allele age. For each trait, we characterized the enrichment of each evolutionary signal compared to matched control regions in the genome that account for linkage disequilibrium, allele frequency, and other confounders. This approach allows us to simultaneously assess the prevalence of multiple evolutionary forces on loci that underlie individual traits as well as between sets of loci that underlie distinct domains of traits. In my presentation, I will describe which forces are detected most frequently across all trait GWASs and within GWASs of specific trait domains. For example, on a subset of high-confidence GWAS we detect pervasive (42/44 traits) and statistically significant enrichments for sequence conservation at trait-associated loci across species and within human populations. I will also discuss several traits whose loci show unique patterns of evolutionary forces and how these results revise our thinking on the evolution of these traits and, more generally, of complex traits.

PrgmNr 3364 - Core promoter similarity in ancient and modern human and in chimpanzee genomes

View session detail

Author Block: **J. L. Dannemiller**; Rice Univ, Houston, TX
Disclosure Block: **J.L. Dannemiller:** None.
Background: Similarity between ancient and modern human DNA sequences has proved useful in understanding the evolutionary history of *Homo sapiens* including major events like *the Out-of-Africa (OoA)* migration. Modern DNA and aDNA sequence similarity probably differs systematically across functional elements within the genome (e.g., exons, promoter sequences, etc.). This study focused on the core promoter region (-50 to +49 bp re: TSS) in Reference and Ancient human genomes (Homo sapiens, *Hs*) as well as in the Chimpanzee (Pan troglodytes, *Pt*) genome. ***Methods:*** The ancient human genome was one of those published by Lipson et al. (2020): a child buried approximately 8000 years ago from the Shum Laka (Cameroon) archeological site. The Reference genome was assembly GRCh38, and the chimpanzee genome was Clint_PTRv2. A novel *follow-on* procedure was used to locate corresponding 100 bp sequences in the Modern and Ancient *Hs* genomes and orthologous sequences in *Pt*. One hundred bp sequences from -150 to -51 bp re: TSS from the Reference genome were used as targets; the Ancient *Hs* and the *Pt* genomes were searched for these targets. A match was declared if there were no more than 2 mismatched nucleotides for a given sequence. A *follow-on* region was then determined by shifting immediately 100 bp downstream: the region from -50 to + 49 bp re: TSS. These aDNA *Hs* and *Pt* core sequences were then compared to the corresponding Reference sequences. There is no purely statistical reason why a match from -150 to -51 bp re: TSS should yield a good match 100 bp downstream. Linkage Disequilibrium, would, of course, predict these to be good matches, but the point is to avoid computing similarity statistics using the target regions because of the biasing effects of the matching requirement. ***Results:*** The Chimpanzee core sequences were slightly better matches to the Modern Reference sequences than they were to the Ancient promoter sequences. Of a total of 893,800 sites compared, there were 363 additional single nucleotide differences between the Ancient and Chimpanzee genomes *vis-Ã -vis* the Chimpanzee/Reference comparison. When conditioning on a nucleotide difference between The Ancient and Modern sequences at a core promoter position, the Chimpanzee nucleotide at that position was more likely to match the Reference allele (67%) than the Ancient allele (29%). ***Conclusions:*** One explanation for these results is that the greater genetic diversity of African genomes relative to non-African genomes (e.g., the Reference) makes the latter more similar to the Chimpanzee outgroup genome. Another way to state this is that a population bottleneck is equivalent to moving a population backward in time toward the MRCA.

PrgmNr 3365 - Estimating average height among populations from polygenic risk scores

View session detail

Author Block: **B. E. Graham**[1], B. Plotkin[2], L. Muglia[3], J. H. Moore[4], S. M. Williams[2,5]; [1]Systems Biology and Bioinformatics, Case Western Reserve Univ., Cleveland, OH, [2]Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [3]Burroughs Wellcome Fund, Research Triangle Park, NC, [4]Dept. of Biostatistics, Epidemiology, & Informatics, Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [5]Genetics and Genome Sci., Case Western Reserve Univ., Cleveland, OH

Disclosure Block: **B.E. Graham:** None.

The genetic basis of phenotypic variation across populations has not been well explained for most traits. Several factors may cause variability in incidence or prevalence, from differences in environments to divergent population genetic structure. We hypothesized that a population level polygenic risk score (PRS), based on common risk allele frequencies, can explain a substantial portion of phenotypic variation among geographic populations. We applied our population specific PRS (psPRS) to 26 populations from the 1000 Genomes to the phenotype height. For each population, we built a psPRS using the 4208 variants shown to associate with height in large GWASs. The psPRSs, in a linear regression model, explained a significant proportion of trait variance for height in men (r^2 = 0.32, p = 0.0026), but less in women (r^2 = 0.11 in women, which was not statistically significant). As not all variants in a PRS may confer similar, or even any, risk among diverse populations, we also performed a filtering algorithm to assess whether variance explained could be improved using psPRSs with fewer SNPs. Variance explained improved with fewer SNPs in the psPRS and was 0.99 for height in men, using only 548 of the initial 4208 SNPs. In women, the SNP pruning method yielded a set of 188 variants with a strong r^2 value of 0.98. The reduced psPRS models were highly significant in both sexes (p -16). That reducing the number of SNPs improves psPRS performance may indicate that missing heritability is partially due to complex architecture that is not additive, the existence of undiscovered population specific variants, or spurious associations in the current databases. We demonstrated that PRS-based analyses can be used across diverse populations for prevalence prediction when applied appropriately, and that these comparisons may be useful in identifying universal risk variants.

PrgmNr 3366 - FastRecomb: Fast inference of genetic recombination rates in biobank scale data using positional Burrows Wheeler transform

View session detail

Author Block: **A. Naseri**[1], W. Yue[1], S. Zhang[2], D. Zhi[3]; [1]Sch. of BioMed. Informatics, Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Univ. of Central Florida, Orlando, FL, [3]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX

Disclosure Block: **A. Naseri:** None.

An accurate genetic map, i.e. estimation of recombination rates along a chromosome is a foundation for genetic studies, including gene mapping, population genetics, and genealogical studies. Recombination rates are known to vary among populations, and thus estimation of population-specific genetic maps is important for advancing genetic research into diverse populations. A traditional approach to infer the recombination rates is to use genotype data from a large number of parent-offspring pairs to capture sufficient numbers of meiotic crossover events. However, collecting parent-offspring pairs can be a practical bottleneck. Another approach involves using population samples to compute the average rates of recombination events by implicitly considering the most recent common ancestor. Current methods using any of the approaches are not computationally efficient enough to handle biobank-scale cohorts comprising millions of individuals. As a result, a subset of the samples has to be selected which translates to losing useful information for inferring recombination rates. Here, we present a novel method using positional Burrows-Wheeler transform data structures and algorithms to identify potential recombination breakpoints in very large cohorts efficiently. The more individuals are available in a data set the more likely the calculated recombination rates are closer to the actual values. Since our method can be applied to samples of millions of individuals without requiring extensive computational resources, we expect that it will enable an accurate and population-specific estimation of recombination rates in diverse populations with the increasing availability of genetic data in large-scale cohorts.

PrgmNr 3367 - Flexible, scalable inference with tree sequences and Pyro

View session detail

Author Block: **W. Wohns**[1], Y. Wong[2], B. M. Neale[3], E. Bingham[1], F. Obermeyer[1]; [1]Broad Inst., Cambridge, MA, [2]Big Data Inst., Li Ka Shing Ctr. for Hlth.Information and Discovery, Univ. of Oxford, Oxford, United Kingdom, [3]Massachusetts Gen. Hosp., Boston, MA
Disclosure Block: **W. Wohns:** None.
Tree sequences encode the genealogical history of a sample of genomes and have been increasingly used as a platform for population genetic inference. However, existing approaches are based on traditional probabilistic methods that are optimized for specific problems and cannot scale to large datasets. Recent advances in deep probabilistic modelling and black box variational inference offer a more flexible approach. Here, we present a simple framework for tree sequence-based inference implemented in Pyro, a deep universal probabilistic programming language (Bingham et al. 2020). We demonstrate this approach using stochastic variational inference algorithms implemented in Pyro to jointly infer the age of coalescent events, past population size, mutation rates, and ancestral geographic location as continuous variables. We evaluate the accuracy of this method using coalescent and forward-time simulations, as well as with tree sequence topologies inferred with tsinfer (Kelleher et al. 2019). Our results are compared with methods including tsdate (Wohns et al. 2021) and Relate (Speidel et al. 2019). Where comparisons exist, our methods provide accuracy comparable with or superior to the state of the art, while scaling to the largest available datasets such as the UK Biobank.

PrgmNr 3368 - Guaranteeing unbiasedness in selection tests based on polygenic scores

View session detail

Author Block: **J. Blanc**, J. J. Berg; Univ. of Chicago, Chicago, IL
Disclosure Block: **J. Blanc:** None.
Population stratification is a well-studied problem in genome-wide association studies, leading to biases in the estimated strength of phenotypic association for individual genetic variants. In short, if environmental effects on the phenotype are correlated with ancestry gradients within a GWAS panel, any variant that is stratified along this ancestry gradient will receive a biased effect size estimate. While state of the art methods to correct for stratification are generally effective in reducing the number of significant false positive associations, even subtle biases in effect size estimates can accumulate across loci, leading to systematic biases in polygenic scores. In turn, these biases in the distribution of polygenic scores can lead to false positives in downstream analyses, such as tests for polygenic adaptation or other analyses of among group genetic differences. One approach is to attempt to be overly aggressive in controlling for stratification. However, there is currently no way to tell conclusively if confounding effects have been removed. A second approach is to conduct the GWAS in an evolutionarily diverged sample that is less likely to share population genetic structure with the test panel. This renders potential biases in the effect sizes irrelevant to the test, but comes at the cost of significantly reduced statistical power due to the issue of poor portability of polygenic scores across samples of divergent ancestry. Here using theory from population and statistical genetics, together with simulations, we show how this second approach can be generalized to panels that do share genetic structure, and that it is possible to guarantee the unbiasedness of selection tests without needing to guarantee that the effect sizes are fully unbiased. Specifically, if the researchers performing the GWAS also have access to the panel of test individuals and have identified the specific test to be performed ahead of time, then it is possible to compute a covariate to include in the GWAS, which will guarantee that the test is unbiased. Further, even when the test is not known prior to conducting the GWAS, our theoretical results provide a way to put a lower bound on the amount of stratification that would be needed to generate a positive signal of polygenic adaptation, and to assess whether the set of principal components that were included in the GWAS are sufficient to render the polygenic adaptation test unbiased. More generally, our results have implications beyond tests for selection as any analysis that attempts to quantify the correlation between polygenic scores and demographic or environmental variables is subject to the same type of stratification biases.

PrgmNr 3369 - GWAS on birth year infant mortality rates provides new evidence of recent natural selection

View session detail

Author Block: **Y. Wu**[1,2], S. Furuya[3], Z. Wang[4], J. E. Nobles[2,3], J. M. Fletcher[2,3,5], Q. Lu[1,2,4]; [1]Dept. of Biostatistics and Med. Informatics, Univ. of Wisconsinâ□□Madison, Madison, WI, [2]Ctr. for Demography of Hlth.and Aging, Univ. of Wisconsinâ□□Madison, Madison, WI, [3]Dept. of Sociology, Univ. of Wisconsinâ□□Madison, Madison, WI, [4]Dept. of Statistics, Univ. of Wisconsinâ□□Madison, Madison, WI, [5]La Follette Sch. of Publ. Affairs, Univ. of Wisconsinâ□□Madison, Madison, WI
Disclosure Block: **Y. Wu:** None.
Recent work on natural selection explores relationships between molecular genetic measurements and realized fitness in the next generation, highlighting the still-evolving nature of human populations. We advance this field by taking a novel approach that examines which genetic variants â□□disappearâ□□ as mortality exposure increases. Specifically, we deploy a â□□regional GWASâ□□ (N = 320,784) that links the county-level infant mortality rate (IMR) by place and year in the UK with genetic variants among birth cohorts in the UK Biobank. The IMR is a well-known indicator of the harshness of the environment during gestation and infancy. These cohorts (born between 1936-1970) saw a decline in IMR from above 65 to under 20 deaths per 1,000 live births, with substantial subnational variations and spikes in 1941 during wartime exposures. We identify two genome-wide significant loci *LCT* and *TLR1/6/10*, both of which are well-known selection targets in Europeans. The beneficial alleles at both loci are associated with higher IMR which is consistent with positive selection in tougher environments. Through analyses stratified in each birth year cohort, we found that effect sizes for both loci peak in 1941, demonstrating accelerated selection on these specific alleles in the UK during the mortality conditions caused by the World War II bombing campaigns during the Blitz. Significant genetic correlations are found across multiple domains, reinforcing earlier findings related to fertility, anthropometrics, and cognition and also extending findings into psychiatric conditions and health conditions. Our study directly gives a genome-wide tests on the shift in allele frequencies due to environmental change and provides fundamental new insights into the mechanism and timing of natural selection in a contemporary European population.

PrgmNr 3370 - Increasing Portability of Predictive Gene Expression Models Across Ancestral Populations Using Genetic Variant Fixation Indices

View session detail

Author Block: **R. Sale**; Vanderbilt Univ., Nashville, TN
Disclosure Block: **R. Sale:** None.
The advent of the 1000 Genomes Project has identified estimations of allele frequencies across ancestral populations for several genetic variants. Differences in allele frequencies across ancestral populations can create confounding problems when attempting to detect potentially causal genetic variants for disease such as in genome wide association studies (GWAS). Furthermore, several studies have also sought to interrogate how gene expression may vary across ancestral populations and have revealed that genetics plays an integral role in gene expression differences across ancestral populations. The ability to detect disease associated variants and, by extension, the ability to impute gene expression by leveraging the genetically determined component of gene expression using PrediXcan remains limited in non-European ancestral populations. In order to improve the ability to assess disease risk and impute gene expression levels for disease phenotypes in non-European ancestral populations, we generate polygenic risk scores (PRS) and gene expression prediction models using SNPs determined to be either cross-population or population differentiated. We calculate fixation index for variants from the 1000 Genomes project to determine whether a variant can be categorized as cross-population or population differentiated. SNPs determined to be cross-population can be utilized to build cross-population PRS and cross-population gene imputation models. These models may have decreased ability to predict disease risk and gene expression levels in European populations but have increased ability to predict disease risk and gene expression levels in non-European populations.

PrgmNr 3372 - The genetic diversity reduction due to background selection increases the disease prevalence under the liability threshold model

View session detail

Author Block: **X. Li**[1], J. J. Berg[2], J. P. Novembre[3]; [1]Committee on Genetics, Genomics & Systems Biology, Univ. of Chicago, Chicago, IL, [2]Univ. of Chicago, Chicago, IL, [3]Univ Chicago, Chicago, IL
Disclosure Block: **X. Li:** None.
How different evolutionary processes maintain phenotypic variation is an important question in human genetics. While the importance of background selection in shaping patterns of neutral genetic diversity is well-studied, its influence on the genetic architecture and the prevalence of complex disease is not well understood.
To address this, we extended the liability threshold model to include background selection. In the model, mutational pressure increases the liability while selection acts to reduce it, and the equilibrium disease prevalence arises due to a balance between the two forces. We found that when background selection is included, it alters this equilibrium by reducing the genetic variance for liability, which drives an increase in disease prevalence. To validate our theoretical results, we performed forward-time SLIM simulations and measure disease prevalence with varying intensity of background selection. We found scenarios where, for example, when the diversity reduction due to background selection is roughly 20% (as estimated in humans), the disease prevalence increases 10%.
Background selection also distorts the allele frequency spectrum at linked loci, as rare variants are less impacted.
To examine whether the distortion of the site frequency spectrum of liability sites changes the disease prevalence, we used simulation to vary the number of rare variants while controlling the genetic diversity reduction. Under the scenarios we examined, the degree of distortion of the site frequency spectrum due to background selection has no meaningful effect on disease prevalence.
From these investigations, we conclude that background selection can impact disease prevalence and does so primarily through overall levels of diversity and not its effect on the relative abundance of rare versus common variants.

PrgmNr 3375 - Addressing the Direct-to-Consumer Genetic Testing Knowledge Gap for Non-Genetics Healthcare Professionals

View session detail

Author Block: **H. Ayoubieh**[1], K. Blazer[2], R. Mills[3], K. Garber[4], H. Lee[5], R. gammal[6], L. Ho[7], K. M. Hyland[8], K. Jacoby Morris[9], M. B. Massart[10], E. Flowers[8], S. M. Robbins[11], G. Kuo[12], D. M. Christopher[11], C. Gunter[13], C. Formea[14], M. Taylor[15], D. Messersmith[16], T. A. Weiler[17], ISCC-PEG DTC-GT Working Group; [1]Texas Tech Hlth.Sci. El paso, El Paso, TX, [2]Director, Cancer Genomics Ed. Program, City of Hope, Duarte, CA, [3]Univ. of North Carolina, Greensboro, NC, [4]Emory Univ Sch. of Med., Atlanta, GA, [5]Park Nicollet Frauenshuh Cancer Ctr., St. Louis Park, MN, [6]MCPHS Univ. Sch. of Pharmacy, Boston, MA, [7]Natl. Ctr. for Advancing Translational Sci., Bethesda, MD, [8]Univ California San Francisco, San Francisco, CA, [9]Natl. Human Genome Res. Inst., BETHESDA, MD, [10]Univ of Pittsburgh, Pittsburgh, PA, [11]Baltimore, MD, [12]UCSD, San Diego, CA, [13]NIH, Bethesda, MD, [14]Univ. of Utah, Utah, UT, [15]Univ. of Colorado, Denver, TX, [16]NHGRI, Bethesda, MD, [17]FIU - HWCOM, Miami, FL

Disclosure Block: **H. Ayoubieh:** None.

Direct-to-consumer genetic testing (DTC-GT) is a convenient method for people to obtain genetic information. Over nine million individuals have used DTC GT services in the U.S., and primary care and specialty physicians are now encountering patients with questions regarding their DTC-GT results. In a recent survey of Kaiser Permanente physicians, 35% of respondents reported that patients had shared their DTC-GT results with them in the past year. Despite the growing need to manage these patient questions, only 12% of surveyed primary care physicians agree that physicians have sufficient knowledge to help patients understand the results of DTC-GT. In a survey of consumers who shared DTC-GT results with their doctors, 27% disagreed with the idea that their physician understood genetics enough to advise them on the implications of their DTC-GT results. Due to this perceived lack of knowledge, many healthcare professionals are reluctant to tackle genomic medicine in their clinical practice and would seek opportunities to enhance their genomic knowledge. The National Human Genome Research Instituteâ□□s Inter-Society Coordinating Committee for Practitioner Education in Genomics (ISCC-PEG)/ / DTC-GT Project Group consists of a spectrum of genetics professionals who aim to improve the genomic literacy of healthcare professionals and enhance the effective practice of clinical genomic medicine in the area of DTC-GT. The ISCC-PEG DTC-GT Project Group has created a Direct-to-Consumer Genetic Testing for Healthcare Professionals Frequently Asked Questions (FAQ) to be housed on Genome.gov. Non-genetics healthcare professionals were surveyed about the FAQ. They found the FAQ content useful, but remarked that it would be difficult to navigate in real time during a clinical encounter. In addition to the FAQ, the DTC-GT Project Group has participated in the 2021 NHGRI Healthcare Provider Genomics Education Week, presenting an educational webinar and short YouTube videos about DTC-GT. The Project Group is also developing a point-of-care tool for healthcare professionals to facilitate an outpatient encounter pertaining to DTC-GT . The Project Group then created 11 DTC-GT patient scenarios, each of which were used to validate the DTC-GT point-of-care tool. The point-of-care tool has now been implemented in the Qualtrics software, as a proof of concept, with the intention to develop a mobile app/website that can support healthcare professionals as they respond to their patients queries about DTC-GT results. Products and outcomes from the DTC-GT Project Groupâ□□s work will be presented.

PrgmNr 3376 - BorderDNA Resources Project: Strategies for the development and disseminationof educational materials on nonmedical DNA applications in immigration contexts

View session detail

Author Block: **N. Mendoza**[1], A. Larson[2], D. J. Madden[3], S. H. Katsanis[4], Z. Guzman[5]; [1]Oberlin Coll., Oberlin, OH, [2]Northwestern Univ., Evanston, IL, [3]Ann & Robert H. Lurie Children's Hosp. of Chicago, Chicago, IL, [4]Lurie Children's Hosp., Chicago, IL, [5]Anne & Robert H. Lurie Children's Hosp., Chicago, IL
Disclosure Block: **N. Mendoza:** None.
DNA testing and collection are increasing a part of U.S. immigration policies, particularly at the border with Mexico. Confusion on the part of stakeholders and the public around the overlapping uses of DNA data are common. Mis- and disinformation are evident in news coverage and social media discussion on these topics, which could be fueled by political differences on immigration. The BorderDNA Resources Project is an academic working group formed to provide unbiased information on DNA uses in U.S. border contexts. We create and disseminate resources on nonmedical DNA applications for journalists, scientists, government officials, migrant families, and the general public. The working group began with pre-identified points of confusion on three specific DNA applications: (A) DNA as used for investigation of transnational missing persons; (B) detainee DNA collection for the federal DNA database CODIS; and (C) DNA tests to verify family relationship claims at the border. We created a master menu to guide the development of products and outreach strategies for each point including audience, product type (infographic, pamphlet, or video), outlet, and language (English, Spanish, or bilingual). We identified eight products on the three topics, including animated videos, infographics, and print materials. We developed product content tailored to the anticipated audience and outlet, keeping in mind sensitivities of the information. To ensure products were unbiased and suitable for distribution, draft products were shared with potentially biased but contrasting representative experts (e.g., NGOs, government, commercial vendors) for feedback. Products were further refined to incorporate feedback. Two finalized infographics were scheduled to be shared three times via the BorderDNA Twitter and Instagram accounts to test our content development and dissemination strategies. Infographics were formatted for publication on each social media platform. Each of the re-posts were altered with variations in captions, direct mentions, and hashtags to test audience engagement. We then assessed the engagement types and levels monitoring audience responses. Early posts showed limited engagement due to the nascent social media account with only one comment outside the study team and a plateau of engagement three days after publication. Additional infographics, print materials, and animated videos are under development. Our strategies to develop and circulate unbiased educational products on nonmedical DNA applications will help inform science communication strategies in other areas with political charge, complex content, and diverse stakeholder groups.

PrgmNr 3377 - Evaluation of genetic services in the care for adults with tetralogy of Fallot

View session detail

Author Block: **S. Estes**[1], L. Wetherill[2], L. W. Markham[3], S. M. Ware[2], B. M. Helm[1], J. S. Needleman[4], M. Sabra[4], J. Herber[3], G. Ephrem[4], K. Spoonamore[4], S. M. Fitzgerald-Butt[2]; [1]Dept. of Med. and Molecular Genetics, Indiana Univ., Indianapolis, IN, [2]Indiana Univ., Indianapolis, IN, [3]Div. of Pediatric Cardiology, Riley Hosp. for Children at Indiana Univ. Hlth., Indiana Univ. Sch. of Med., Dept.s of Pediatrics and Med., Indianapolis, IN, [4]Krannert Inst. of Cardiology, Dept. of Med., Div. of Cardiology, Indiana Univ. Sch. of Med., Indianapolis, IN

Disclosure Block: **S. Estes:** None.

Background: While prior studies have investigated the prevalence of underlying genetic conditions in adult congenital heart disease (ACHD) patients with tetralogy of Fallot (TOF), no studies have sought to determine the extent to which current guidelines for inclusion of genetic services and genetic testing are being implemented. The most common syndromic cause of TOF is 22q11.2 deletion syndrome (DS), and most babies born with TOF are routinely screened for 22q11.2 DS. However, because most genetic evaluations are conducted in infancy, there remains a potential gap in care for adult patients born with TOF prior to routine genetic testing for this condition. As a result, American Heart Association ACHD care guidelines recommend testing for this deletion in patients with conotruncal CHDs, of which TOF is the most common. **Methods:** We conducted a retrospective chart review of ACHD patients with TOF seen in cardiology clinic during a 5-year period, 2015-19, via examination of cardiac and genetics-related clinic documentation available in the electronic health record. We determined the proportion of patients who had genetic services and genetic testing for 22q, and Fisher's exact tests were used to assess for association between these and patient year of birth (YOB), medical history, and clinical factors. A *p*-value Results: Of the 299 patients, 54 (18%) had received genetic services and 18 (6%) had documented testing for 22q11.2 DS. While patient YOB was not associated with genetic services/testing, developmental/cognitive issues, dysmorphic facial features, extracardiac physical anomalies, biological sex, and other health histories (seizures, nasal speech, and/or hearing/vision loss) were associated with genetic services/testing. When the above variables were analyzed together via a logistical regression model, developmental/cognitive issues, dysmorphic facial features, and extracardiac physical anomalies were significantly associated with having received genetic services/testing. **Conclusion:** Despite current guidelines for genetic testing in all patients with conotruncal CHDs, it is clear that traditional indicators of a syndromic presentation still drive the majority of genetic services/testing in this ACHD population. This is an opportunity for expansion of genetic services/testing to this population as understanding the genetic etiology of TOF in adults can significantly impact their medical screening as well as their reproductive/family planning.

PrgmNr 3378 - Exploring the Use of Family Health History in Black Communities in the Deep South

View session detail

Author Block: **C. Whitted**[1], F. Gilpin-Macfoy[1], T. Lee[2], L. Koehly[1]; [1]NHGRI/NIH, Bethesda, MD, [2]Florida A&M Univ., Tallahassee, FL
Disclosure Block: **C. Whitted:** None.
Family health history (FHH) is an underutilized tool with the potential to influence health behavior within the family context. Family health communication channels within Black families may be an existing pathway to influence health behaviors related to diseases with a strong genetic link. The Families SHARE workbook is a tool designed to encourage communication of family health history about diseases that have a familial pattern. The tool is used to calculate risk and provides recommendations for heart disease, type 2 diabetes, colon, breast, and prostate cancer. Black populations living in southern regions of the United States have high incidence and mortality rates for the diseases included in the Families SHARE workbook. Here, we aimed to evaluate usability and acceptability of the Families SHARE workbook and explored the potential use of the tool to increase awareness of FHH in Black adults living in the southern region of the United States. In addition, we examined if the workbook could improve awareness of family genetic risk information and encourage prevention behaviors in Black adults living in the southern region of the United States. A mixed methods approach was used to evaluate associations and outcomes by evaluating overall workbook use and differences in the workbookâ□□s usability, cognitive and behavioral impact in Black adults living in North Florida. A total of 62 participants completed baseline surveys and 49 participants completed follow up surveys. A total of 10 participants completed semi structured interviews on workbook use and disease awareness. Overall, 79% of participants indicated the ability to use the workbook. Workbook usability outcomes show that eighty-five percent of participants were able to able to assess their own degree of risk. Black women were more likely to share the workbook with others (65% vs.20%; p<.05 and were more likely to be able assess a family members degree of risk vs. p compared black men. an independent samples t-test was conducted examine cognitive behavioral outcomes differences in communication after receipt the families share workbook. participants that understood sd="2.351);" t="2.073" communicate about disease this study shows populations living south may effectively use workbook understand fhh information with for chronic diseases. next steps will seek usage combined sample.>

PrgmNr 3379 - Genetic service provision for men with prostate cancer within a safety-net setting: A qualitative study

View session detail

Author Block: **E. Li**[1], C. Wang[2], G. A. Gignac[1], K. Zayhowski[3], M. Pankowska[1], C. M. Gunn[1,2]; [1]Boston Univ. Sch. of Med., Boston, MA, [2]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [3]Boston Med. Ctr., Boston, MA

Disclosure Block: **E. Li:** None.

Recent changes to the Prostate Cancer Early Detection National Comprehensive Cancer Network Guidelines have expanded the criteria for genetic testing, increasing the number of patients who are eligible for cancer genetics services. Data from large cancer centers indicate that only 40-44% of oncologists have provided referrals to genetic germline prostate cancer genetics care (Loeb et al., 2020; Paller et al., 2019). Less is known about the provision of prostate cancer genetics services within safety-net settings, where resource constraints may impact access for patients. Through purposive sampling of providers who facilitated cancer genetics care at a large, urban safety-net institution, we conducted semi-structured telephone interviews with 15 health care providers (physicians, nurse practitioners, physician assistants, and genetic counselors) from medical and radiation oncology, urology, cancer genetics, and primary care. Facilitators and barriers to delivering genetics care were identified at the patient, provider, and health system levels. At the patient level, providers perceived that patientsâ health knowledge, trust, and attitudes towards genetic testing affected uptake of referrals. Provider-level factors focused on the perceived utility of genetics information for clinical care, knowledge of rapidly changing genetics guidelines, and perceived ability of patients to pay for genetics services. Systems-level barriers included the lack of technological infrastructure to identify potentially eligible patients to cancer genetic counseling, a need for increased staff capacity to facilitate patient completion of genetic testing (i.e., genetic counselors and navigators), and lacking provider education to promote genetics care. Provider-identified opportunities to improve genetics care encompassed the need to 1) establish an identification and referral system to cancer genetics by augmenting institutional medical record systems, 2) increase in the number of personnel to support genetic counseling services, and 3) create tools to enhance patient and provider genetics knowledge. Within a safety-net setting, it is important for providers who facilitate germline genetics care for patients with prostate cancer to understand and respond to the unique needs of their patient community and pursue opportunities that mitigate barriers to accessing genetic services.

PrgmNr 3380 - Genetic-first diagnoses and disease predispositions: Implications for follow-up care

View session detail

Author Block: **C. Jodarski**[1], A. Katz[2], M. Kamen[1], M. Setzer[1], J. Yan[1], L. Jamal[3], K. Jevtich[1], Y. Yu[1], B. Seifert[1], R. Ghosh[1], S. Holland[1], M. A. Walkiewicz[1], M. Similuk[1]; [1]Natl. Inst. of Allergy and Infectious Diseases, Bethesda, MD, [2]Natl. Human Genome Res. Inst., Bethesda, MD, [3]Natl. Cancer Inst./NIH Clinical Ctr. Dept. of Bioethics, Bethesda, MD
Disclosure Block: **C. Jodarski:** None.
As genomic testing, including sequencing and microarray, becomes more common in clinical practice, so will receiving a â□□genetic-firstâ□□ result. We define this result as a likely pathogenic or pathogenic variant in a gene that is unrelated to the primary indication for testing, which can include a secondary finding (SF) from the ACMG gene list. Such results offer the opportunity for early intervention, but due to the difficulty of clinical interpretation in seemingly asymptomatic individuals, the value varies by case. Here we present a series of â□□genetic-firstâ□□ cases from our clinical research at the National Institute of Allergy and Infectious Diseases that emphasize the nuances of such findings: Exome sequencing (ES) was performed on a 12-year-old male with *CYBB*-related chronic granulomatous disease, who also had feeding difficulties and ADHD. A likely pathogenic *OTC* variant was detected, which is a cause of OTC deficiency and a possible explanation for some aspects of this patientâ□□s history. Subsequent biochemical testing was normal. The recommended follow up included repeat biochemical testing during active illness but not dietary modification. This case underscores that a SF sometimes only infers a risk state and does not necessarily lead to a clinical diagnosis or active disease mitigation. A 46-year-old female underwent ES as part of a trio-based analysis for her sonâ□□s phenotype. Her medical history was unremarkable, and family history included a sister with melanoma and a father with metastatic lung cancer. ES identified a pathogenic variant in *TP53*, which is associated with Li-Fraumeni syndrome (LFS). Based on her personal and family history, she was unlikely to have been identified as at-risk for LFS. However, due the identification of this SF, she can receive appropriate screening and management and alert at-risk relatives. A 5-year-old female, noted to be â□□clumsyâ□□, presented for disseminated coccidioidomycosis with brain and pulmonary involvement. ES and microarray detected a 1.330 Mb gain at 17p12 consistent with a molecular diagnosis of Charcot-Marie-Tooth disease that was supported by subsequent nerve conduction studies. Without genetic testing, her peripheral neuropathy could have been misinterpreted as a worsening of her infections. Learning of this result early in her disease course allows for proactive management of her neuropathy. These cases highlight opportunities and challenges of â□□genetic-firstâ□□ results. As genomic approaches become better integrated into clinical practice, thoughtful evaluation and tailored follow-up care will be essential in translating such findings into meaningful recognition and mitigation of disease.

PrgmNr 3381 - Grandmothers save the day: The power of grandmothers to improve care of families with history of cardiomyopathy

View session detail

Author Block: **S. M. Fitzgerald-Butt**, K. Spoonamore, B. Helm, K. Schmit, N. Gralia, E. Geddes, S. M. Ware, J. J. Parent; Indiana Univ., Sch. of Med., Indianapolis, IN
Disclosure Block: **S.M. Fitzgerald-Butt:** None.
Obtaining an accurate, targeted family history (FH) is critical to informing high-quality care for patients being evaluated for personal or FH of cardiomyopathy. This can take substantial time, prompting genetic counselors to seek alternative solutions to gather FH information prior to the multidisciplinary clinic appointments including forms (paper or electronic), chatbots, and/or phone calls. While much energy has been devoted to gathering this pre-clinic FH information, for select patients, there is a benefit to gathering FH information post-clinic. We present a series of three pediatric patients for whom obtaining additional FH information from a grandmother following the initial encounter improved the patientâ□□s/familyâ□□s care. The first patient had a known diagnosis of familial dilated cardiomyopathy (DCM) and reported his sister and father, both deceased, had DCM. While we were aware of the sisterâ□□s DCM diagnosis, a discussion with the maternal grandmother surprisingly uncovered the father did not have DCM. Rather, the maternal FH included multiple relatives with history suspicious for DCM, sudden death, cardiomyopathy, and even heart transplant. The second patient had a maternally-inherited pathogenic variant associated with hypertrophic cardiomyopathy (HCM) identified as a secondary finding on exome sequencing and reported no FH of HCM. Upon contacting the maternal grandparents to arrange cascade genetic counseling/testing, the grandmother reported she previously had a mildly abnormal chest x-ray and EKG for which follow-up was not recommended, her brother had died suddenly after being told he had â□□heart problemsâ□□ and suspicious histories in her father and paternal uncle. The grandmother shared the familial variant and her cardiac evaluation revealed HCM. The last family, while knowledgeable of their significant paternal FH of DCM, were unaware that the great-grandfather had completed genetic testing that identified a likely pathogenic variant. Following a discussion with the grandmother, who allowed access to her medical records, we learned of the specific variant for which she had received counseling but not testing. It was discussions with grandmothers which informed screening and genetic testing recommendations for the correct side of the family, identified an affected individual related to a secondary finding, and enabled targeted genetic testing in each of these respective families. Post-clinic collection of FH information, especially from grandparents, is a useful endeavor. Resources to support these efforts are valuable for genetic counselors.

PrgmNr 3382 - Integrating Ethical, Legal, and Social Implications (ELSI) considerations in the *All of Us* Research Program:Experiences of the ELSI Brain Trust

View session detail

Author Block: **S. Chandrasekharan**[1], S. Sodeke[2], R. Isasi[3], L. Beskow[4], J. Reusch[5], B. Berkman[6], K. Blizinsky[1], The All of Us ELSI Brain Trust; [1]The All of Us Res. Program/NIH, Bethesda, MD, [2]Tuskegee Univ., Tuskegee, AL, [3]Univ. of Miami Miller Sch. of Med., Miami, FL, [4]Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]All of Us Res. Program, NIH, Bethesda, MD, [6]NIH, Bethesda, MD
Disclosure Block: **S. Chandrasekharan:** None.
From its inception, the *All of Us* Research Program has looked critically and prospectively at ELSI considerations, starting with its core values and in its stance toward inclusion, diversity, and participant engagement. The program strives to increase geographical and demographic diversity, while committing to repairing and building trust with groups and communities who have suffered historical harms from, or have been under-represented in, biomedical research (UBRs). It also endeavors to provide diverse and rich data--including genomic, health, social, environmental, and lifestyle data--to foster precision medicine research that advances health and well-being for all. The programâ□□s strong focus on inclusion of individuals from racial and ethnic UBR groups, specifically, aims to alter persistent trends in genomic studies that predominantly include individuals of European ancestry and urban, high socio-economic status groups. This ethos of meaningful, longitudinal inclusion combined with collecting multi-dimensional data from highly diverse populations at a national scale, raises familiar and new ELSI, policy, and regulatory issues. These include scenarios where there are variations in policies and best practices among research initiatives and gaps in knowledge pertaining to these diverse populations. Responsively, the program created an internal working group, the ELSI Brain Trust (EBT), to deliberate on ELSI issues in a timely and contextual way, and inform the programâ□□s design, implementation, and policy development. In this presentation we will provide background on the programâ□□s overall approach for conducting precision medicine research in a diverse cohort and returning genetic results in ways that respect participantsâ□□ choices and enable voluntary decision-making. Using examples of EBT consultations, we will describe ELSI considerations that emerge in areas such as (1) inclusion of vulnerable groups (children and those with decisional impairment), (2) protecting participants and preventing harm while making annotated genomic sequence and sensitive data available to a wide range of research users, and (3) evaluating and continually improving the programâ□□s processes for responsible return of results. We will also share insights from using a consultation model that embeds the EBT within the program, and enumerate potential future challenges, particularly ELSI considerations for emerging areas in genomic research (e.g., use of AI/ML and returning polygenic risk scores).

PrgmNr 3383 - Measuring genetic literacy through familiarity, skills, and knowledge: A longitudinal assessment

View session detail

Author Block: **I. Little**[1], L. Koehly[1], C. Gunter[2]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]NIH, Bethesda, MD
Disclosure Block: **I. Little:** None.
As genomic and personalized medicine is integrated into healthcare, the need for patients to understand and make informed decisions about their genetic risk information increases. Genetic literacy, or oneâ□□s knowledge of genetic principles and their applications, measures an individualâ□□s ability to apply genetic information to their life, including their health. Increased genetic literacy can improve comprehension of genetic tests and therefore increase participation in healthcare services to detect and treat genetic disorders. Increased genetic literacy on both sides can also improve patient-provider communication in this new era of precision medicine. However, measures for genetic literacy vary widely and often assess genetic knowledge exclusively. Because current literature indicates that the populationâ□□s genetic literacy is generally low, there is a need to adequately measure genetic literacy to identify potential gaps and education strategies. A 2015 paper adapted previous surveys into a new measure that assesses genetic literacy in three facets aimed to mirror the way individuals learn new concepts in everyday life. We replicated this measure by distributing a comprehensive survey to a nationally representative sample of 2050 individuals, assessing genetic literacy in the same three facets: familiarity with genetic terms (e.g. heredity, sporadic); accuracy in genetic knowledge (e.g. Is a gene bigger than a chromosome?); and sufficient skills to synthesize information applying genetic information to human health (e.g. What is the purpose of genetic testing?). Results indicated that genetic literacy increased over time. Participants reported only moderate familiarity with genetic terms on a 1 - 7 scale (M = 5.28, SD = 1.38) though it was significantly higher than 2015 (p

PrgmNr 3384 - Patient perspective on return of incidental findings in Africa

View session detail

Author Block: **S. Prochaska**[1], A. Oladayo[2], T. Busch[3,4], W. Adeyemo[5], L. Gowans[5], M. A. Eshete[6], W. Awotoye[3,4], A. Alade[4,3], A. A. Adeyemo[7], P. Mossey[8], A. Prince[9], J. C. Murray[10], A. Butali[11,4]; [1]Coll. of Publ. Hlth., Univ. of Iowa, Iowa City, IA, [2]Univ. of Iowa, Iowa City, IA, [3]Dept. of Oral Pathology, Radiology and Med., Coll. of Dentistry, Univ. of Iowa, Iowa City, IA, [4]Iowa Inst. for Oral Hlth.Res., Univ. of Iowa, Iowa City, IA, [5]Komfo Anokye Teaching Hosp. and Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [6]Addis Ababa Univ., Sch. Of Med., Dept. of Surgery, Addis Ababa, Ethiopia, [7]NIH, Bethesda, MD, [8]Dept. of Orthodontics, Univ. of Dundee, Dundee, United Kingdom, [9]Coll. of Law, Univ. of Iowa, Iowa City, IA, [10]Dept. of Pediatrics, Univ. of Iowa, Iowa City, IA, [11]Dept. of Oral Pathology, Radiology and Med., Coll. of Dentistry, Iowa City, IA

Disclosure Block: **S. Prochaska:** None.

Objective: Orofacial clefts (OFCs) are the most common birth defects of the head and neck region, affecting 1 out of 700 live births worldwide. OFCs lead to significant financial, educational, medical, psychological, and cultural problems and may require treatment throughout life. Whole genome sequencing (WGS) is a cost-effective way to identify causal variants across the genome for families with OFC. Our group has generated data from the first WGS effort for OFC in African population. WGS data also provides an opportunity for the identification of secondary findings like sickle cell anemia, G6PD deficiency, and other single gene disorders that may impose medical implications for those in the study or their family members. Secondary findings may impose some burden on families. Therefore, the objective of this study was to determine how to manage incidental findings in a resource-limited setting by determining the interest of OFC families in receiving incidental genetic information and medical risks, including possible identification of infectious disease and the outcomes of those who choose to be informed.**Methods:** In December of 2020, a Qualtrics survey was administered to recruited families from cleft-craniofacial clinics in Ethiopia, Ghana, and Nigeria.**Results:** In the results of the survey as of April 30th, 2021, 109 individuals completed the survey. 93 participants had a household total family income of less than $10,000 (USD). A total of 101 agreed or strongly agreed that they would want to know about secondary findings, 7 did not agree nor disagree, and 1 strongly disagreed. When asked if they felt that there were currently available resources that allowed them to access genetic information that they wanted to know, 10 agreed and 99 disagreed.**Conclusion:** In this pilot, the biggest barrier that we observed was access to genetic information at the OFC clinics in Africa. The validated survey assessed patientâ□□s willingness to receive information regarding secondary findings and available resources in Africa to deal with secondary findings. This is the first ELSI study to document the interest of participants in OFC studies in incidental genetic information and medical risks in an African population.

PrgmNr 3385 - Patients demonstrate high motivation to enhanced cancer screening following genetic counselling

View session detail

Author Block: **K. Stafford**, C. Jani, O. Al Omari, H. Singh, L. Agarwal, A. Dar, M. Lotz, A. D'Angelo, P. B. Lam; Mount Auburn Hosp., Cambridge, MA
Disclosure Block: **K. Stafford:** None.
Background: Early cancer screening has been shown to reduce morbidity and mortality in patients at high risk of cancer. High-risk personal or family history of cancer is usually used as a screening tool to help identify individuals who will benefit from more accurate risk stratification with genetic testing for hereditary cancer syndromes. In this retrospective observational study, we sought to characterize the indications, resulting recommendations, and patient response to the results of genetic counselling.
Methods: We included patients undergoing genetic testing at our hospital between January 2018 and October 2020 (N=237). Medical charts were reviewed to identify patient characteristic, family history, indication for genetic testing, genetic findings, and subsequent management.
Results: Our population had a mean age of 52.5 years (SD=14.8) and was predominantly female (93.4%). The vast majority (195, 82.3%) had a family history of cancer, while 113 (47.7%) had a personal history of cancer, most commonly breast (n=87). Most of the patients met criteria for genetic testing through Hereditary Breast and Ovarian Cancer criteria (n=226, 95.4%). Of the 235 tested, 36 were found to have deleterious mutations (15.3%). The most common resulting recommendation was initiation of yearly breast MRI alongside yearly mammograms (n=100, 42.5%). Other common screening tools implemented included pancreatic cancer screening with MRCP and EUS (n=24, 10%), early initiation of colonoscopy and endoscopy screening (n=20, 8.5%), yearly skin exams (n=10, 4.2%), and initiation of prophylactic hormone therapy for breast cancer (n=10, 4.2%). Several patients had already undergone mastectomy or oophorectomy, however 9 patients (3.8%) were recommended oophorectomy, and 3 (1.2%) were recommended mastectomies. Sixty-two patients (26.4%) were recommended to continue ongoing oncology care while 37 (15.7%) were advised to continue with PCP follow-up and routine screening. Overall, recommendations were well received with 97.5% (n=231) of patients stating intent to follow-through with recommendations.
Conclusions: Enhanced cancer screening recommendations were well received by patients following consultation in a high-risk cancer clinic, with 97.5% intending to follow-through with cancer screening recommendations. Among patients who underwent genetic cancer testing, 15.3% were found to have a pathogenic mutation.

PrgmNr 3386 - Recruitment of diverse participants into a trial of polygenic risk scoring: Enrollment metrics and reasons for decline in the Genomic Medicine at VA (GenoVA) Study

View session detail

Author Block: **A. Antwi**[1], N. Jones[1,2], N. Alexander[1], C. Brunette[1], E. Harris[1,2], J. L. Vassy[1,2,3]; [1]VA Boston Hlth.care System, Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Brigham and Women's Hosp., Boston, MA

Disclosure Block: **A. Antwi:** None.

Background: People of non-European ancestry continue to be underrepresented in genomic research despite significant advancements made over the past two decades. Improving the diversity of participants in genomic medicine research is vital to ensuring existing health disparities are not further exacerbated. Here, we present the early results of an enhanced-diversity recruitment strategy for the GenoVA Study.

Methods: The GenoVA Study is a randomized trial of polygenic risk scoring (PRS) for 6 common diseases. Eligible participants are patients of the VA Boston Healthcare System (VABHS) aged 50-70 years without prior diagnosis of atrial fibrillation, coronary artery disease, type 2 diabetes, colorectal, breast, or prostate cancers. Study staff identify potentially eligible participants via electronic health record data query, mail recruitment letters with an opt-out option, confirm study eligibility and obtain informed consent by phone. Non-white, Hispanic, and female participants are oversampled for recruitment to increase enrollment of underrepresented populations at VABHS. Study staff collect reasons for declining participation during the initial phone call. Consented participants provide a biospecimen for PRS calculation and interpretation from a CLIA-certified laboratory.

Results: The GenoVA study aims to enroll 1,074 participants. As of June 7, 2021, a total of 10,501 patients across VABHS were deemed eligible for study participation. Of these, 3040 patients have been contacted, 2797 patients were reached by phone, 2333 did not opt out of recontact, and 395 consented to participation. Of the first 382 enrolled patients, mean (SD) age at enrollment is 59.5 (5.5) years, 32% are women, and 38% are Hispanic or non-white, compared to 12% of women and 14% of Hispanic or non-white patients in the eligible VABHS population overall. Among the 542 participants who actively declined participation, common reasons included having no interest (n=202), time constraints (n=113), health reasons (n=44), disliking research studies (n=28), personal reasons (n=25), not wanting to learn about their genetic information (n=23), privacy or security concerns (n=19), VA or government-specific reasons (n=19), and emotional distress or anxiety about receiving genetic results (n=12).

Discussion: The characteristics of the first GenoVA Study enrollees demonstrate the success of an enhanced-diversity recruitment strategy. As clinical PRS results are now being returned to participants, this greater representation of underserved populations will help to expand the equitable reach of genomic medicine research.

PrgmNr 3388 - Should laboratories report small regions of homozygosity?

View session detail

Author Block: **N. Urraca**[1], S. Holtrop[2], R. Mostafavi[3], L. Holly[1], K. Bosfield[1], J. C. Ward[1], C. W. Brown[1], E. K. Pivnick[1]; [1]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [2]Le Bonheur Children's Hosp., Memphis, TN, [3]St. Jude Children's Res. Hosp., Memphis, TN

Disclosure Block: **N. Urraca:** None.

SNP-array is often used as a first tier test in clinical genetics evaluations and may incidentally identify regions of homozygosity (ROH). Reporting thresholds differ among labs, with most reporting ROH â¥5 Mb. Multiple ROH may suggest identity by descent (IBD) and increased risk for autosomal recessive (AR) disorders in offspring, while a single ROH could be due to segmental uniparental disomy (UPD) or distant consanguinity. While autozygosity mapping in all cases of ROH can aid in identifying genes responsible for disease, genetic counseling for UPD differs from counseling of consanguineous couples regarding potential psychosocial effects, risk of AR conditions in offspring and risk of recurrence. We present 3 cases in which SNP-array initially reported a single ROH >5 Mb, but multiple ROH of 2-5 Mb were found upon closer examination. **Case 1**: 2-day-old with multiple congenital anomalies. SNP-array revealed a 7.67 Mb ROH on chromosome 17. Further analysis detected multiple ROH Case 2: 21 month old with global developmental delay, coarse facies, tachypnea and abdominal distension. MPS urine and enzyme studies were non-diagnostic. SNP-array reported a 5.74 Mb ROH at 1p36.12, which included *FUCA1,* the gene associated with fucosidosis. Parents were from the same small city but denied consanguinity. Upon request, the clinical lab identified a total of 50 Mb of smaller ROH, indicating IBD. **Case 3**: newborn with an extended cystic lymphangioma. A 6 Mb ROH on chromosome 16 was initially identified. Four additional ROH

PrgmNr 3389 - Strategies to increase clinicians in medical genetics: Establishing a pipeline to attract and prepare diverse clinical researchers

View session detail

Author Block: **D. Murray**[1], S. D. Fernbach[2], L. C. Burrage[1], R. Sutton[1], B. Lee[1]; [1]Baylor Coll. Med., Houston, TX, [2]Baylor Coll. of Med., Houston, TX
Disclosure Block: **D. Murray:** None.
In 2014, analyzing Association of American Medical Colleges (AAMC) data, the NIH Physician-Scientist Workforce Working Group reported that within the last few decades, there had been a steady decline in physicians conducting research 5.5% since 2003, along with a lack of diversity in the physician-scientist workforce. The medical genetics residentsâ data in the Molecular & Human Genetics (MHG) department indicates a minimal number of underrepresented groups in the Internal Medicine and Pediatric Genetics tracks. These data underscore the critical need to attract more underrepresented in medicine (UiM) trainees to the scientific and clinical enterprise. In the past three years, the MHG department focused attention on attracting and preparing underrepresented groups to develop and provide innovative training principles to provide UiM researchers the knowledge and ability to apply genetics and genomics to a range of healthcare conditions paving the way to increase health equity. We devised a recruiting strategy to attract and prepare UiM interested in genetics from three sources: (1) those found in our learning environment (summer interns), (2) college campuses, and (3) medical schools. In 2019, a Lunch and Learn seminar was created to introduce genetics and genomics careers to the undergraduate summer students training in MHG laboratories. The following summer, we created the virtual Genetics and Genomics Careers (GGC) series to reach undergraduates around the country and introduce them to these careers. From the three webinars, the following data were collected: 72 students attended, from 4 Texas high schools, 23 U.S. universities, one each from the United States Air Force (USAF) and M.D. Anderson Cancer Center. In addition, we offered a townhall to medical students around the country to learn and discuss the medical genetics pathway. The virtual townhall attracted 54 medical student attendees, with survey respondents from 10 states. Lastly, we wanted to reach first-year medical students to introduce the medical genetics career and recruit for a new summer research internship. In 2021, we visited 1st-year medical students at Meharry Medical College and did outreach activities at the University of Texas - Rio Grande Valley. We recruited two students to the Clinical Research Education Training Program. Employing multiple recruiting strategies demonstrated that underrepresented undergraduate and medical students have an interest in genetics and genomics careers.

PrgmNr 3390 - The Texome Project: Special considerations for developing a rare and undiagnosed disease program for underserved communities in Texas

View session detail

Author Block: **T. Nguyen Dolphyn**[1,2], C. A. Bacino[1,3,4], C. M. Eng[1,4], P. Liu[1,4], Z. Liu[1,2], S. Yamamoto[1,2,5,6], M. F. Wangler[1,2,5], H. J. Bellen[1,2,5,7]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Jan and Dan Duncan Neurological Res. Inst., Texas Childrenâ□□s Hosp., Houston, TX, [3]Texas Children's Hosp., Houston, TX, [4]Baylor Genetics Lab., Houston, TX, [5]Program in Dev.al Biology, Baylor Coll. of Med., Houston, TX, [6]Dept. of NeuroSci., Baylor Coll. of Med., Houston, TX, [7]Howard Hughes Med. Inst., Houston, TX

Disclosure Block: **T. Nguyen Dolphyn:** None.

The Undiagnosed Diseases Network (UDN) has exemplified the role genomic medicine can play in solving medical mysteries and providing families with much needed answers. Due to a number of factors, however, underserved populations and ethnic minorities have not always benefited as fully from large-scale genomics research and the benefits of genomic medicine in general. The Texome Project is a new research program leveraging the study teamâ□□s experience and success in clinical evaluations, sequencing, bioinformatics, and functional studies to implement and expand the reach of personalized genomics and medicine to diverse populations within Texas. Specifically, the program aims to help minority and underserved individuals with low socioeconomic status who have an undiagnosed, suspected genetic disease. To reach a socioeconomically and ethnically diverse cohort, patients and families are recruited through a multi-faceted strategy that includes community outreach and education, partnership with diverse physicians and local community hospitals, and a systematic chart review within the Texas Childrenâ□□s Hospital system to identify patients with public insurance (e.g. Medicaid) and genetic testing denial. Rather than requiring extensive genetics workup or medical record review as criteria for study enrollment, this program reduces barriers by interfacing with providers and families directly to focus on diagnostic needs. Embedded within the study design is a streamlined process to facilitate physical examination, DNA collection, genetic testing and counseling, and return of results, all with telemedicine options to ease logistical or transportation burden for patients. Longitudinal follow-up appointments over 2 years provides patients who remain undiagnosed with access to further bioinformatics analyses and genomic match-making. The follow-up also allows for examination of patient outcomes, medical management, and patient and family attitudes towards genomic medicine whether they receive a diagnosis or not. We have developed a protocol for the Texome Project that is sensitively designed to collect reliable outcome data to explore and understand the barriers and facilitators of implementing genomic and precision medicine in a diverse population. We anticipate this project will highlight the need and value of including underserved and underrepresented patients to further diagnosis and gene discovery in rare disease.

PrgmNr 3392 - Uptake of cancer risk management strategies among women who undergo cascade genetic testing for breast cancer susceptibility: How do they compare to non-cascade testers?

View session detail

Author Block: **S. Makhnoon**[1], L. Brooke[2], K. Mattie[2], R. Volk[1], G. Grana[3], B. Arun[1], S. K. Peterson[1]; [1]The Univ. of Texas MD Anderson Cancer Ctr., Houston, TX, [2]MD Anderson Cancer Ctr. at Cooper Univ. Hlth.Care, Camden, NJ, [3]Cooper Hlth System, Voorhees, NJ

Purpose: A primary rationale for cascade testing for hereditary breast and ovarian cancers (HBOC) is to offer cancer risk management options, including bilateral mastectomy (BLM), salpingo-oophorectomy (BSO), and intensified screening, to persons found to have an inherited predisposition to the disease. At-risk relatives of probands with a pathogenic variant have familial experiences with cancer genetic testing which may shape health beliefs and subsequent health behaviors. Yet, little is known about their genetically-informed cancer risk-reduction behaviors, and long-term outcome data among cascade testers are lacking. This study evaluated: (1) the uptake and timing of cancer preventive surgical and screening strategies between cascade and non-cascade testers, and (2) the association between uptake of cancer risk management strategies and proband characteristics. Methods: Medical records were abstracted for all unaffected women with pathogenic variants in HBOC-associated genes from two cancer hospitals with at least one year of follow-up to compare uptake of surgery and screening between cascade and non-cascade testers. Results: 341 women underwent post-test genetic counseling between 2013 and 2019, of whom 253 were included in the analytic sample. Cascade testers (79.8%) were younger than non-cascade testers (mean=37.6 vs. 43.5 years, p=0.002) and most commonly had a parent with a pathogenic variant (39.1%) followed by siblings (21.3%), and multiple other relatives (20.8%). Women were predominantly non-Hispanic White (81.0%) and underwent testing for *BRCA1* (42.0%) or *BRCA2* (47.2%) variants. Among women ageâ¥40 years, 43% underwent BLM and 71.6% underwent BSO with no significant difference in uptake between cascade and non-cascade testers. Mean time to BSO among cascade testers was shorter among women ageâ¥40 vs Conclusion: Management uptake among cascade testers is high with rates comparable to unaffected *BRCA* positive women. A large proportion of women act on cascade test results and this represents a novel report of utilization of cancer management strategies

PrgmNr 3393 - Colchicine: A Potential Therapy for the HSPB8-related Rimmed Vacuole Myopathy

View session detail

Author Block: **J. Vu**[1], L. Weiss[2], A. Shmara[3], E. Mohanty[4], R. Yan[5], R. Wei[5], C. Cheng[6], V. E. Kimonis[7]; [1]Univ. of California Irvine, Sch. of Med., Dept. Pediatrics, Orange, CA, [2]Univ. of California, Irvine, [3]UCI, Yorba Linda, CA, [4]Univ. of California, Irvine, Orange, CA, [5]Univ. of California Irvine, Irvine, CA, [6]UC Irvine, Irvine, CA, [7]Univ CA Irvine, Orange, CA
Disclosure Block: **J. Vu:** None.
Our lab has recently reported a family with a novel disorder associated with autosomal dominant Rimmed Vacuole Myopathy (RMV) caused by a frameshift mutation: c.515dupC in the Heat Shock Protein family B member 8 (HSPB8) gene. HSPB8 is a chaperone involved in the Chaperone Assisted Selective Autophagy (CASA) complex. HSPB8 in conjunction with BAG3 recognizes and promotes the autophagy-mediated removal of misfolded proteins present in several neurodegenerative diseases. Patients with this debilitating disease exhibit proximal limb-girdle with distal myopathy, muscle atrophy, fatty replacement, and rimmed vacuoles. Our previous studies show that patients with HSPB8 c.515dupC have reduced expression of HSPB8 protein, disrupted TDP-43, and autophagy pathology in patient fibroblasts. Colchicine, after an extensive drug screening project in neuronal cells, has shown to be one of the top drugs to significantly increase expression of HSPB8 as well as block misfolded TDP-43 accumulation. We hypothesize that treating transgenic HSPB8 mutant mice with Colchicine will help to restore normal HSPB8 gene function and block TDP-43 accumulation/mislocalization. Thus, improving motor function and endurance of the HSPB8 mutant mice. We treated our newly created c.515dupC HSPB8 knock-in mouse model with colchicine (0.3 mg/kg/day for 3 months). The dose was derived from a previously conducted dose-response study. After treatment, we have observed (1) significant improvement of muscle pathology by H&E staining (2) increased HSPB8 RNA and protein levels, and (3) a trend of amelioration in rotarod performance in the heterozygous treated group compared with the vehicle group. The results from this study indicate that colchicine has the potential for treating HSPB8 associated rimmed vacuole myopathy.

PrgmNr 3394 - Genome-wide integration assay for rAAV mediated homologous recombination (HR) in human hepatocytes demonstrated precision of *in vivo* gene editing approach

View session detail

Author Block: **J. F. Thompson**, J. Von Stetina, J. Prout, M. Tian, A. Pla, S. Potts, M. Rubin, R. Resendes, G. Cerqueira, J. Wright; Homology Med.s, Bedford, MA
Disclosure Block: **J.F. Thompson:** Salary/Employment; Homology Medicines Inc.
There have been many studies examining the integration of DNA arising from AAV and other viruses into sites throughout the human genome. Most current methods use amplification of sequences on both sides of the integration junction. Typically, specific primers directed to the virus being studied are combined with primers directed at non-specific adaptors ligated to genomic DNA near the junction. However, a variety of artifacts can obscure real integration signals, and these are often caused by a high degree of background amplification or from unintended ligation of free or truncated virus to random genomic DNA during processing. This is particularly a problem with assessing AAV-mediated HR integration sites because, at high doses, episomal AAV may be present at thousands of copies per cell and thus easily amplified or ligated inadvertently. False positive integration signals arising from the very high episomal background may occur more frequently than actual integration events. These issues have led to questions regarding the authenticity of some reported genomic integrations.
To overcome the high episomal background often observed with AAVs, and other known sources of artifactual â□□integrationsâ□□, we have modified previous protocols and combined them with new steps to eliminate the impact of excess vector genomes. Long-read sequencing has been employed to ensure that relevant viral and human genomic sequences are truly present on the same molecule, eliminating index hopping as an issue as seen with short-read sequencing. The mean length of molecules spanning the integration junction that we sequenced is approximately 2 kb and some molecules are over 6 kb. A positive control cell line has been generated with a single inserted DNA via CRISPR. Genomic DNA from this cell line is spiked at known fractional concentrations into genomic DNA from human hepatocytes engrafted in mice so that the frequency of real events can be estimated. These mice were treated with rAAV with a codon-optimized *PAH* gene and homology arms targeted for insertion into intron 1 of human *PAH*. Using these quantitative molecular methods, we are able to calculate the frequency of HR-mediated integration into the desired locus. We see no evidence of integration into any other genomic location, which further supports the precision of this gene editing approach.

PrgmNr 3395 - A New Paradigm for Coordinate-Based, Multipanel Genomic Plotting in R

View session detail

Author Block: **N. Kramer**[1], E. S. Davis[1], C. Wenger[2], S. Parker[1], D. Phanstiel[1]; [1]UNC Chapel Hill, Chapel Hill, NC, [2]Independent Scholar, Issaquah, WA
Disclosure Block: **N. Kramer:** None.
The R programming language is unparalleled in its ability to transform raw genomic data sets into meaningful biological conclusions through analysis and visualization. However, when it comes to complex, multi-paneled plots, users rely on 3rd party graphic design software to arrange and customize plots. Here we present BentoBox, a coordinate-based genomic data visualization package that introduces a new paradigm in multi-figure plotting in R. Employing grid Graphics and drawing from the practices of base plotting and ggplot2, BentoBox offers precise, programmatic control over the placement, aesthetics, and arrangements of plots through a user-defined page with explicit units of measurement. Its edge-to-edge, containerized visualizations preserve the mapping between user-specified containers and the represented data, ensuring accurate alignment of plots in the same genomic region. Operating entirely within the R environment and tightly integrated with Bioconductor, BentoBox is optimized to quickly and easily read, plot, and arrange multi-omic data for an endless number of use cases. Come join us in changing how we plot in R!

PrgmNr 3396 - All of Us Researcher Workbench: Bringing Researchers to Genomics Data

View session detail

Author Block: **A. Bick**[1], A. Able[1], A. Musick[2], E. Banks[3], M. Basford[1], P. Harris[1], L. Lichtenstein[3], K. Mayo[1], B. Ozenberger[2], A. Roman[1], S. Rura[4], D. M. Roden[5], All of Us Data and Research Center; [1]Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]NIH, Rockville, MD, [3]Broad Inst., Cambridge, MA, [4]Verily Life Sci., Cambridge, MA, [5]Vanderbilt Univ. Med. Ctr.,, Nashville, TN
Disclosure Block: **A. Bick:** None.
The NIHâ□□s All of Us Research Program[1] aims to collect health data from a million or more participants to create a diverse research resource that accelerates precision medicine. The All of Us Researcher Workbench[2] is a secure, cloud-based platform developed by the Data and Research Center (DRC) through which registered researchers can access and analyze data from program participants including data from electronic health records, surveys, physical measurements, and wearables from ~315,000 participants. The DRC has developed new tools that provide efficient, scalable, and low-cost solutions for processing and managing these data using Google BigQuery as a serverless, cloud-native solution. Advantages of the cloud-based workbench design include enhanced security, speed, scalability and decreased cost.

All of Us researchers can generate project-specific Workspaces, select and annotate cohorts of the All of Us dataset, and export analysis-ready tables to interactive Jupyter Notebooks for analysis. To date >860 researchers from >215 institutions have leveraged this free-to-access platform.

In 2020, All of Us Genome Centers began generating both array genotyping and whole genome sequence (WGS) data from participant biosamples. A critical next milestone towards accomplishing All of Usâ□□ mission is providing analysis tools and a large diverse genomic dataset in the All of Us Researcher Workbench. Within the Workbench, researchers will have the ability to select
All of Us participants that have WGS and/or genotype array data and perform large-scale analyses using standard tools such as Hail or Plink. We anticipate enabling access to the genomics capabilities to all users of the All of Us Researcher Workbench by the end of 2021, including ~80,000 WGS and ~120,000 genotype arrays. Importantly, we anticipate that ~44% of participants with WGS data will be from racial and ethnic communities that have been historically underrepresented in biomedical research.

To ensure genomics community usability of this new â□□bringing researchers to the genomic dataâ□□ paradigm, the All of Us Research Program Science Committee has launched four genomics demonstration projects: Variant Frequency, GWAS, Genetic Ancestry, and Array Polygenic Risk Score. By design these demonstration projects do not aim to make new discoveries. Rather, they aim to validate the All of Us cohort genomics data and tools by replicating known, previously published findings.

(1) All of Us Research Program Investigators. (2019). The â□□All of Usâ□□ research program. New England Journal of Medicine, 381(7),
668-676.(2)https://www.researchallofus.org/data-tools/workbench/

PrgmNr 3397 - Annotation of genetic variation in 5134 cystic fibrosis cases with OpenCRAVAT and VEP

View session detail

Author Block: **K. A. Pagel**[1], H. Nebeck[2], D. Jain[3], M. Aksit[1], H. Ling[1], R. Pace[4], K. Raraigh[5], F. Onchiri[2], A. Faino[2], A. Stilp[2], E. E. Blue[2], F. A. Wright[6], M. J. Bamshad[2], Y-H. Zhou[7], R. Gibson[2], M. R. Knowles[4], R. Karchin[1], G. R. Cutting[1], S. M. Blackman[1]; [1]Johns Hopkins Univ., Baltimore, MD, [2]Univ. of Washington, Seattle, WA, [3]Seattle, WA, [4]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [5]Johns Hopkins Univ., Columbia, MD, [6]North Carolina State Univ, Raleigh, NC, [7]North Carolina State Univ., Raleigh, NC
Disclosure Block: **K.A. Pagel:** None.
Introduction: In-depth analysis of variation from large scale sequencing projects requires accurate and comprehensive annotation of variant properties and application of filters to reduce the search space. Here, we describe the approaches used in the processing and annotation of variation from large scale whole genome sequencing data in a cohort of 5134 cases with cystic fibrosis and strategies to address common pitfalls. Methods: We performed the annotation for each chromosome separately, with the genotype annotations removed to reduce runtime and memory consumption. We used vt to perform variant decomposition and left-normalization to split multi-allelic variants and ensure accurate representation of insertion/deletion variants. We required two annotation tools to provide the curated annotations that are needed to perform rare variant aggregation analysis with variant filtering strategies specified by TOPMed. First, we applied VEP v101 to obtain 5kb upstream and LOFTEE annotations. Next, we used OpenCRAVAT to incorporate annotations for Ensembl Regulatory Build, promoters from GeneHancer and several prediction scores. Annotation source data were obtained from dbNSFP where available, and the primary sources otherwise. Results: Variants that impact more than one gene were annotated separately for each gene, with the MANE canonical transcript used to identify transcript-specific features when available. We applied the TOPMed variant specifications and generated filtered variant aggregate files, with a novel script to assess the characteristics of each variant per transcript. We describe further variant-level characteristics that should be used in curation of the analysis set including minor allele count, allele frequency, and quality. We identify and address multiple factors that impact the quality and accuracy of annotation tools including canonical transcript selection, changes in gene definitions and characterization of multi-allelic variation. Conclusions: We believe that the variant pre-processing and annotation pipeline constructed in this work can serve as a guide on the critical factors to ensure accurate annotation for variation in large scale sequencing projects.This study is supported by CFF grants CUTTIN18XX1, BAMSHA18XX0, and KNOWLE18XX0 and is submitted on behalf of the CF Genome Project.

PrgmNr 3398 - FusionGDB 2.0, fusion gene annotation update aided by deep learning

View session detail

Author Block: **P. Kim**[1], H. Tan[1], J. Liu[1], H. Lee[2], H. Jung[3], H. Kumar[1], X. Zhou[1]; [1]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Korea Inst. of Oriental Med., Daejeon, Korea, Republic of, [3]Asan Med. Ctr., Seoul, Korea, Republic of

Disclosure Block: **P. Kim:** None.

Gene fusion is one of the hallmarks of the cancer genome via chromosomal rearrangement initiated by DNA double-strand breakage. A knowledge base with the systematic functional annotation of fusion genes is critical for understanding genomic breakage context and developing therapeutic strategies. FusionGDB is a unique functional annotation database of human fusion genes and has been widely used for studies with diverse aims. In this study, we report FusionGDB 2.0, which has substantial updates of contents such as up-to-date human fusion genes, fusion gene breakage tendency score with FusionAI deep learning model based on 20kb genomic sequence around BP area, investigation of overlapping between fusion breakpoints with 44 human genomic features across five cellular roleâ□□s categories, transcribed chimeric sequence and following open reading frame analysis with coding potential based on deep learning approach with Ribo-seq read features, and rigorous investigation of the protein feature retention of individual fusion partner genes in the protein level. Among ~ 126k fusion genes, about 30k kept their ORF as In-frames, which is triple size compared to the previous version, FusionGDB. FusionGDB 2.0 will be used as the reference knowledge base of fusion gene annotations. FusionGDB 2.0 provides seven categories of annotations: Fusion Gene Summary, Fusion Gene ORF, Fusion Gene Genomic Feature, Fusion Protein Feature, Fusion Sequence, Related Drug, and Related Disease.

PrgmNr 3400 - Genomics4RD: An integrated platform to share Canadian deep-phenotype and variant level data for international rare disease gene discovery

View session detail

Author Block: **H. Driver**[1], T. Hartley[1], M. Price[2], O. Buske[3], A. K. Ramani[4], M. Osmond[5], K. D. Kernohan[2], Care4Rare Canada, B. M. Knoppers[6], M. Brudno[7], K. M. Boycott[8]; [1]Children's Hosp. of Eastern Ontario Res. Inst., Ottawa, ON, Canada, [2]CHEO, Ottawa, ON, Canada, [3]Gene42, Toronto, ON, Canada, [4]Hosp for Sick Children, Toronto, ON, Canada, [5]Ottawa, ON, Canada, [6]McGill Univ., Westmount, QC, Canada, [7]Hosp Sick Children/Univ Toronto, Toronto, ON, Canada, [8]Children's Hosp. of Eastern Ontario, Ottawa, ON, Canada

Disclosure Block: **H. Driver:** None.

The low incidence of some rare genetic diseases (RDs) means that data-sets for gene discovery need to be as large as possible; sharing of information between research groups is therefore critical. The Care4Rare Canada Consortium (previously FORGE Canada) is a pan-Canadian network of clinicians and researchers that has been studying undiagnosed RDs for over a decade. In 2018, the consortium launched a 4 year project called C4R-SOLVE, a sub-aim of which was to collect, harmonize, and share both retrospective and prospective Canadian phenotypic and multiâomic data. Here, we introduce Genomics4RD; an integrated web-accessible platform to share Canadian phenotypic and variant-level genetic data between researchers both within Canada and internationally, for the purpose of rare disease gene discovery.

Genomics4RD has been designed to facilitate precision medicine for patients with RDs, by addressing current and projected barriers to data sharing; including concerns about privacy, discrepant study protocols, jurisdiction-specific regulations, social concerns amongst users, and low resources. At the core of Genomics4RD, is an unambiguously structured governance framework that dictates how participant data is stored, tagged (using ADA-M codes), accessed, and withdrawn. This governance framework is linked to a library of consent templates and REB waivers to aid data contributors in releasing data from institutional custodianship. Data storage, authorization, and access procedures have been developed in collaboration with stakeholders to ensure the trusted and secure access of data by external researchers. In collaboration with PhenoTips, the database has been designed to aid users in systematically collecting, prioritizing, and visualizing the phenotypic and genotypic information of participants. All raw multi-omic sequencing data is processed through centralized bioinformatics pipelines, to produce compatible datasets for analyses. The breadth and consistency of data offered by Genomics4RD allows researchers to compare putative disease genes across patients (i.e. single-sided matching), greatly increasing the possibility of disease gene discovery.

Genomics4RD offers modern solutions to combat barriers to data sharing, and allows for international collaborations in RD research. Our platform minimizes risks for patients, clinicians, and researchers, and facilitates greater diagnostic yields for RDs.

PrgmNr 3401 - HGRepo - A comprehensive repository of clinical, lab assay, and genetic data for studies of innate and adaptive immune responses during acute COVID-19 infection and convalescence

View session detail

Author Block: **J. Barnett**[1], S. Samuel[1], A. J. Oler[1], Z. Li[1], X. Cheng[1], C. Shyu[1], M. Nambiar[1], K. Kaladi[1], V. Kuram[2], Y. Zhang[3], S. Weber[4], W. Lau[5], M. Stack[6], M. Magliocco[4], E. Shaw[3], M. S. Abers[3], O. M. Delmonte[3], G. Montealegre[3], H. Su[3], L. Notarangelo[3], S. Xirasagar[1]; [1]Bioinformatics and Computational BioSci.s Branch, Office of Cyber Infrastructure and Computational Biology, Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [2]NIAID Collaborative Bioinformatics Resource, Frederick Natl. Lab. for Cancer Res., NIH, Frederick, MD, [3]Lab. of Clinical Immunology and Microbiol., Div. of Intramural Res., Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [4]Molecular Dev. of the Immune System Section, Lab. of Immune System Biology, and Clinical Genomics Program, Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [5]NIH Ctr. for Human Immunology, Multiscale Systems Biology Section, Lab. of Immune System Biology, Natl. Inst. of Allergy and Infectious Diseases, NIH, Bethesda, MD, [6]Lab. of Clinical Immunology and Microbiol., Natl. Inst. for Allergy and Infectious Diseases, NIH, Bethesda, MD
Disclosure Block: **J. Barnett:** None.

We describe a unique clinical omics portal (HGRepo) developed to enable collaborative access to longitudinal clinical events and multiple types of biomarker phenotypic data to provide comprehensive profiles of the disease progression in COVID-19 patients. It enabled harmonized representation of data for the NIAID COVID-19 Immune Response international collaboration, which was spearheaded in March 2020 to unveil the innate and adaptive immune responses resulting from SARS-CoV2 infection. The synergistic coalition of > 15 researchers began collaborative analysis of the standardized and curated COVID-19 data in the portal within three months of receipt and published early results in August 2020. Developed on LabKey, HGRepo provides intuitive interfaces for data loading and curation, data export (spreadsheets and API) system administration (e.g., defining tables, role-based user accessioning), browsing, and querying. Iterative implementation helped support evolving consortium needs. Automated data transformation to standardize incoming data in heterogeneous formats from multiple (>15) international sites is performed in close collaboration with clinicians and informatics researchers. The harmonized representation of data is intuitive to clinicians and is also amenable for data science approaches to analysis. Chronological representation of clinical events (such as COVID-19 tests, COVID symptoms, hospitalization) and sample collection and linking of samples to assay data (WGS, RNAseq, serological markers, etc.) enabled users to efficiently and accurately discover relationships between the disease progression and genetics and omics data. The portal also provides access to summarized genetic data derived from Whole Genome Sequencing (WGS), and research serology data. The portal is being used to identify biomarkers as surrogates of disease progression as well as in clinical classification of the disease. As of June 2021, HGRepo includes 6810 subjects and 23671 samples from over 15 sites. Computational processing and analysis of more than 1600 WGS combined with the clinical data in HGRepo have contributed to the discovery of novel variants in immune-related genes that correlate with COVID-19 disease severity. Immunological and virological correlates and predictors of clinical outcomes have been published in five articles to-date. Importantly, these data can be combined with other future studies. We are now using this collaborative framework of development to systematically store standardized data for other NIAID labs and are confident in our ability to rapidly stand-up scientific data management systems for future pandemic studies.

PrgmNr 3402 - New genomic data deposited in dbGaP cross-referenced against NINDS Repository Tourette Syndrome biospecimens

View session detail

Author Block: **L. B. Scheinfeldt**[1], S. Heil[1], A. Green[1], A. Amberson[1], S. Sander-Effron[1], R. Zhang[2]; [1]CORIELL Inst. FOR Med. RESEARCH, Camden, NJ, [2]NIH/NINDS, Rockville, MD
Disclosure Block: **L.B. Scheinfeldt:** None.
The National Institute of Neurological Disorders and Stroke (NINDS) Human Genetics Resource Center was established by the NINDS as a public biorepository of DNA and lymphoblastoid cell lines collected from subjects diagnosed with neurological disorders as well as population controls. Each biospecimen is annotated with associated de-identified clinical and demographic data. Since the establishment of the NINDS Repository, specimens from over 42,000 individuals have been successfully banked at the Coriell Institute for Medical Research and can be accessed through an online catalog (http://catalog.coriell.org/NINDS). The collection includes samples from individuals diagnosed with cerebrovascular diseases (N>12,700), Parkinsonism (N>5,600), motor neuron diseases (N>2,500), Epilepsy (N>6,000), Tourette syndrome (N>4,200), Dystonia (N>3,500), and neurologically normal controls (N>7,500). To date, NINDS Repository samples have been used in over 570 peer-reviewed scientific publications, including genome-wide association studies, whole-exome sequencing studies, and structural variation studies. The genomic data collected by these studies have been deposited in the Database of Genotypes and Phenotypes (dbGaP), a NIH/NLM sponsored restricted-access data repository for studies investigating the contributions of genetic variation to phenotypic variation and disease (http://www.ncbi.nlm.nih.gov/gap). The NINDS Repository catalog includes a resource page devoted to the annotation of genomic data collected from NINDS Repository samples that are available in dbGaP (https://www.coriell.org/1/NINDS/NINDS-Samples-in-dbGaP). Recently, this resource page was updated to include two large-scale dbGaP genomic datasets: one that includes genome-wide SNP data generated from biospecimens donated by 1,880 Tourette Syndrome patients, and another which includes whole exome sequencing data collected from biospecimens donated by 378 Tourette Syndrome patients. An overview of this annotated information can be accessed via a downloadable spreadsheet that details each biospecimen associated with genomic data in a dbGaP study accession, and the corresponding study accession details. This dbGaP annotation resource for the NINDS Repository ensures that existing genomic data collected from Repository samples is easily associated with its original biospecimen and made available for reuse in new genomic studies of Tourette Syndrome, neurological disorders, and human disease more generally.

PrgmNr 3403 - A Finnish-enriched functional DNA variant in an enhancer for *IRF6* is associated with cleft palate

View session detail

Author Block: **F. Rahimov**[1], P. Nieminen[2], P. Kumari[3], E. Juuri[4], T. Nikopensius[5], A. Palotie[6], FinnGen, A. HeliÃ¶vaara[7], W. D. Fakhouri[8], B. C. Schutte[9], R. A. Cornell[3], D. Rice[4]; [1]Genomics Res. Ctr., AbbVie, North Chicago, IL, [2]Univ. of Helsinki, Helsinki, Finland, [3]Dept. of Anatomy and Cell Biology, Univ. of Iowa, Iowa City, IA, [4]Helsinki Univ. Hosp., Univ. of Helsinki, Helsinki, Finland, [5]Estonian Genome Ctr., Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [6]Inst. for Molecular Med. Finland, Helsinki Inst. of Life Sci., Univ. of Helsinki, Helsinki, Finland, [7]Cleft Palate and Craniofacial Ctr., Dept. of Plastic Surgery, Helsinki Univ. Hosp., Helsinki Univ., Helsinki, Finland, [8]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [9]Dept. of Microbiol. and Molecular Genetics, Michigan State Univ., East Lansing, MI
Disclosure Block: **F. Rahimov:** Salary/Employment; AbbVie.

Genetic variation in the Interferon Regulatory Factor 6 (*IRF6*) gene causes and contributes risk for orofacial clefting. Here we performed a genome-wide association study for orofacial clefting in the Finnish population and identified a significant association with rs570516915, a single nucleotide polymorphism that changes a highly conserved nucleotide in a known enhancer for *IRF6*, called MCS-9.7. We found that the risk allele diminishes the activity of the enhancer using reporter assays, and CRISPR-Cas9 edited oral epithelial cells demonstrate rs570516915 decreases expression of *IRF6*. Previous associations at the *IRF6* locus were predominantly observed with cleft lip (CL) or cleft lip with cleft palate (CLP). However, the association with this DNA variant was driven solely by cases of isolated cleft palate (CP) (p -34, OR = 8.65, 95% CI 6.11-12.25). We replicated the association in an independent sample of CP cases from Estonia (p -11). This finding is significant to public health because unlike other countries, the frequency of CP in Finland is higher than the frequency of CL or CLP. Moreover, the frequency of the risk allele for rs570516915 parallels the regional variation of CP prevalence in Finland (p IRF6 harboring rs570516915 is required for proper development of both the palate and the lip, two embryologically distinct processes.

PrgmNr 3404 - Analysis of genetic investigations in newborns with VACTERL spectrum of malformations

View session detail

Author Block: **I. Chacon**[1], A. James[2], R. Mendoza[2], D. Chitayat[3]; [1]Lakeridge Hlth.Oshawa, Oshawa, ON, Canada, [2]Hosp. for Sick Children, Toronto, ON, Canada, [3]Ontario Power Generation Bldg,, 700 Univ. Ave Rm 3-709, Mount Sinai Hosp., Toronto, Ontario,, Toronto, ON, Canada
Disclosure Block: **I. Chacon:** None.
VACTERL association is an acronym that stands for **v**ertebral defects, **a**nal atresia, **c**ardiac defects, esophageal atresia and **t**racheo-**e**sophageal fistula, **r**enal and **l**imb abnormalities. It has a non-random occurrence, usually with a good prognosis following surgical management and with low-recurrence risk. It is a diagnosis by exclusion; however, many environmental, maternal, chromosomal and single gene disorders mimic the same pattern of malformations and therefore a minimum genetic diagnostic protocol should be undertaken to rule out other diagnosis. We developed a genetic diagnostic flow chart for newborns with VACTERL spectrum of malformations, based on the evidence for test diagnostic yield. We reviewed all admissions to our Hospitalâ□□s Neonatal Intensive Care Unit from 2010 to 2018 with three or more of the VACTERL core features. The most common genetic investigation was microarray analysis (68%), followed by chromosome analysis. Taken together, all patients had one of these tests as the first or only test done. The second most common test was DNA analysis for suspected single gene disorder (27.3%), followed by chromosomal breakage studies (20.2%) and whole exome sequencing (13%). Overall, the diagnostic yield of genetic testing was low (4.6%). Only in 6% of the patients a diagnosis was obtained following genetic testing including: Trisomy 18 (2), Cat Eye syndrome (1), Coffin-Siris syndrome (1) and Feingold syndrome (1). The two most informative genetic testing were FISH and chromosome analysis. Microarray analysis had no diagnostic yield (provided no further diagnostic information) with the incidence of VUS being almost 20%. Based on our data analysis and review of the literature, our flow chart recommends that in cases suspected of having VACTERL association who have no facial dysmorphism, CNS malformation, no limb involvement or features of a particular genetic syndrome, genetic evaluation could be deferred to 10-12 months at which time, if no new issues (ex. failure to thrive or developmental delay), no further genetic test is required.

PrgmNr 3405 - Biallelic variants in *KARS1* are associated with neurodevelopmental disorders and hearing loss recapitulated by the knockout zebrafish

View session detail

Author Block: **G. K. Varshney**[1], S-J. Lin[2], B. Vona[3], P. G. Barbalho[4], R. Kaiyrzhanov[5], R. Maroofian[5], C. Petree[4], M. Severino[6], V. Stanley[7], P. Varshney[1], P. Bahena[8], F. Alzahrani[9], A. Alhashem[10], A. Pagnamenta[11], G. Aubertin[12], J. Estrada-Vera[13], H. Adrian Diaz-Hernandez[14], N. Mazahery[15], A. Oza[16], J. Thies[17], R. Deborah[18], S. Dugad[19], J. McEvoy[20], T. Sultan[21], L. Pais[22], B. Tabarki[10], D. Villalobos-Ramirez[23], A. Rad[3], Genomics England Research Consortium, H. Galehdari[24], F. Ashrafzadeh[25], A. Sahebzamani[26], K. Saeidi[27], E. Torti[28], S. Mora[28], T. Palculict[28], H. Yang[28], M. Joshi[19], M. Behra[29], S. Burgess[30], S. K. Nath[4], M. Hanna[5], M. Kenna[31], H. Houlden[5], E. Ghayoor[32], M. S. Zaki[33], T. H. Haaf[34], F. S. Alkuraya[35], J. Gleeson[20]; [1]Genes & Human Disease Res. Program, Oklahoma Med. Res. Fndn., Oklahoma City, OK, [2]Oklahoma Med. Res. Fndn., Oklahoma CIty, OK, [3]Tubingen Hearing Res. Ctr., Eberhard Karls Univ. of Tubingen, Tubingen, Germany, [4]Oklahoma Med. Res. Fndn., Oklahoma City, OK, [5]Univ. Coll. London, London, United Kingdom, [6]Neuroradiology Unit, IRCCS Istituto Giannina Gaslini, Genoa, Italy, [7]Univ. of California San Diego, La Jolla, La Jolla, CA, [8]Julius Maximilians Univ. Wurzburg, Wurzburg, Germany, [9]KFSHRC, Riyadh, Saudi Arabia, [10]Prince Sultan Military Med. City, Riyadh, Saudi Arabia, [11]Univ. of Oxford, Oxford, United Kingdom, [12]Island Hlth., Victoria Gen. Hosp.,, Victoria, BC, Canada, [13]Henry M Jackson Fndn. for the Advancement of Military Med., Bethesda, MD, [14]Natl. Inst. of Med. Sci. and Nutrition Salvador Zubira, Mexico City, Mexico, [15]Shahid Chamran Univ. of Ahvaz, Ahvaz, Iran, Islamic Republic of, [16]Harvard Med. Sch., Boston, MA, [17]Seattle Childrens Hosp., Seattle, WA, [18]Mayo Clinic Coll. of Med. and Sci., Rochester, MN, [19]Bioinformatics Ctr., S. P. Pune Univ., Pune, India, [20]Univ. of California San Diego, La Jolla, CA, [21]Children's Hosp. and Inst. of Child Hlth., Lahore, Pakistan, [22]Broad Inst., Boston, MA, [23]Univ. of Wurzburg, Germany, Wurzburg, Germany, [24]Dept.s of Genetics, Shahid Chamran Univ., Ahvaz, Iran, Islamic Republic of, [25]Pediatric and Genetic Counselling Ctr., Kerman Welfare Organization, Kerman, Iran, Islamic Republic of, [26]Pediatric and Genetic Counselling Ctr., Kerman, Iran, Islamic Republic of, [27]Kerman Univ. of Med. Sci., Kerman, Iran, Islamic Republic of, [28]GeneDx Inc., Gaithersburg, MD, [29]Dept. of Anatomy and Neurobiology, Sch. of Med. UPR, San Juan, PR, [30]NHGRI, NIH, Bethesda, MD, [31]Boston Children's Hosp., and Dept. of Otolaryngology, Harvard Med. Sch., Boston, MA, [32]Univ. of London, London, United Kingdom, [33]Natl. Res. Ctr., Dokki, Cairo, Egypt., Egypt, [34]Julius Maximilians Univ. of Wuerzburg, Wuerzburg, Germany, [35]King Faisal Specialist Hosp. & Res. Ctr., Riyadh, Saudi Arabia

Disclosure Block: **G.K. Varshney:** None.

Pathogenic variants in Lysyl-tRNA synthetase 1 (KARS1) have increasingly been recognized as a cause of early-onset complex neurological phenotypes. To advance the timely diagnosis of KARS1-related disorders, we sought to delineate its phenotype and generate a disease model to understand its function in vivo. Through international collaboration, we identified 22 affected individuals from 16 unrelated families harboring biallelic KARS1 likely pathogenic or pathogenic variants. Sequencing approaches ranged from disease-specific panels to genome sequencing. We generated loss of function alleles in zebrafish. We identify ten new and four known biallelic missense variants in *KARS1* presenting with a moderate-to-severe developmental delay, progressive neurological and neurosensory abnormalities, and variable white matter involvement. We describe novel KARS1-associated signs such as autism, hyperactive behavior, pontine hypoplasia, and cerebellar atrophy with prevalent vermian involvement. Loss of *kars1* leads to upregulation of p53, tissue-specific apoptosis, and downregulation of neurodevelopmental related genes, recapitulating key tissue-specific disease phenotypes of patients. Inhibition of p53 rescued several defects of *kars1*$^{-/-}$ knockouts. Our work delineates the clinical spectrum associated with KARS1 defects and provides a novel animal

model for KARS1-related human diseases revealing p53 signaling components as potential therapeutic targets.

PrgmNr 3406 - Congenital heart defects caused by *FOXJ1*

View session detail

Author Block: **M. Padua**[1], B. Helm[2,3], J. Wells[2], A. Smith[1], H. Bellchambers[1], A. Sridhar[4], S. M. Ware[1,2]; [1]Dept. of Pediatrics, Indiana Univ., Indianapolis, IN, [2]Dept. of Med. & Molecular Genetics, Indiana Univ., Indianapolis, IN, [3]Dept. of Epidemiology, Indiana Univ., Indianapolis, IN, [4]Indiana Univ., Indianapolis, IN
Disclosure Block: **M. Padua:** None.
Heterozygous pathogenic variants in the Forkhead box J1 (FOXJ1) transcription factor cause a motile ciliopathy leading to situs inversus, obstructive hydrocephalus, and chronic airway disease. *FOXJ1* is expressed in ciliated cells of the airways, testis, oviduct, central nervous system, and the embryonic left/right organizer. Prior studies indicate ablation or targeted mutation of *Foxj1* in mice, zebrafish and frogs results in a reduced number of motile cilia, which affects the establishment of the left-right body asymmetry. However, congenital heart defects (CHD) resulting from loss of function of this transcription factor have not been well characterized in model organisms or humans. We performed phenotypic characterization of *Foxj1* null embryonic 14.5-17.5 hearts in mice revealing either normal (levocardia; 57%) or abnormal heart looping (43%) such as dextrocardia or sinistral looping (24.6%), ventral looping (heart tube loops forward in the frontal plane;16% of which nearly half displayed a combined ventral/sinistral looping) and no looping or single ventricle heart (2.4%). Further histological analysis showed a wide range of cardiac defects including atrial, ventricular, and atrioventricular septal defects, non-compacted ventricular walls as well as double outlet right ventricle defects. Given the range of CHD findings in mouse, we analyzed a cohort of 279 participants with CHD phenotypes associated with left/right defects using research-based exome sequencing and identified three individuals with predicted damaging missense variants in *FOXJ1*. Subsequently, we identified a hospitalized patient using clinical exome sequencing who had a novel mutation in exon 3 of *FOXJ1* (*c.784_799dup*; p.Glu267Glyfs*12). The patientâ□□s CHD included double outlet right ventricle, ventricular septal defect, trabecular hypertrophy of the right ventricle, and mild left ventricular outflow obstruction. Currently, experiments in *Xenopus laevis* embryos are ongoing for functional analysis of these *FOXJ1* variants. These experiments use the known role of *FOXJ1* as a master regulator of ciliogenesis to assess the ability of the identified variants to form ectopic cilia in the epidermis at developmental stage 26, similar to wildtype *FOXJ1*. These results indicate that in addition to its role as a master ciliogenesis gene, *FOXJ1* plays an important role in cardiogenesis and abnormalities in *FOXJ1* can cause CHD.

PrgmNr 3407 - Loss of *Zic3* impairs planar cell polarity leading to abnormal left-right signaling and heart defects

View session detail

Author Block: **H. Bellchambers**[1], S. M. Ware[1,2]; [1]Dept. of Pediatrics, Indiana Univ. Sch. of Med., Indianapolis, IN, [2]Dept. of Med. & Molecular Genetics, Indiana Univ. Sch. of Med., Indianapolis, IN
Disclosure Block: **H. Bellchambers:** None.
Loss of function of *ZIC3* causes heterotaxy, a disorder characterized by organ laterality defects including complex heart defects. Studies using *Zic3* mutant mice have demonstrated that loss of *Zic3* causes heterotaxy due to defects in establishment of left-right (LR) signaling, but the mechanistic basis for these defects remains unknown. Here, we demonstrate *Zic3* null mice undergo cilia positioning defects at the embryonic node consistent with impaired planar cell polarity (PCP). Cell-based assays demonstrate that ZIC3 must enter the nucleus to regulate PCP and identify multiple critical ZIC3 domains required for regulation of PCP signaling. Furthermore, we show that *Zic3* displays a genetic interaction with the PCP membrane protein *Vangl2* and the PCP effector genes *Rac1* and *Daam1* resulting in increased frequency and severity of neural tube and heart defects. Gene and protein expression analyses indicate that *Zic3* null embryos display disrupted expression of PCP components and reduced phosphorylation of the core PCP protein DVL2 at the time of LR axis determination. These results demonstrate that ZIC3 interacts with PCP signaling during early development, identifying a novel role for this transcription factor, and adding additional evidence about the importance of PCP function for normal LR patterning and subsequent heart development.

PrgmNr 3408 - Novel pathogenic variants, evidence for involvement of *DVL2*, and quantitative phenotypic analyses of Robinow syndrome: WNT signaling perturbation and phenotypic variability

View session detail

Author Block: **C. Zhang**[1], A. Jolly[1,2], B. J. Shayota[3,1,4], J. F. Mazzeu[5,6], H. Du[1], M. Dawood[1,2,7], P. C. Soper[8], A. de Lima[5], B. M. Ferreira[5], Z. Coban-Akdemir[1,9], J. J. White[1], D. Shears[10], F. R. Thomson[10], S. L. Douglas[11], A. Wainwright[10], K. Bailey[10], P. Wordsworth[12], M. Oldridge[10], T. Lester[10], A. D. Calder[13], K. Dumic[14], S. Banka[15], D. Donnai[16], S. N. Jhangiani[7], L. Potocki[17], W. K. Chung[18], S. Mora[8], H. Northrup[9], M. Ashfaq[9], J. A. Rosenfeld[1], K. Mason[19,8], L. C. Pollack[19], A. McConkie-Rosell[20], W. Kelly[20], M. T. McDonald[20], N. S. Hauser[21], P. Leahy[22], C. M. Powell[23], R. Boy[24], R. S. Honjo[25], D. M. Muzny[7], R. A. Gibbs[26,7], J. E. Posey[1], P. Liu[1,27], J. R. Lupski[1,4,7], R. Sutton[26,4], C. M. B. Carvalho[28,1]; [1]Baylor Coll. of Med., Houston, TX, [2]Med. Scientist Training Program, Houston, TX, [3]Univ. of Utah, Salt Lake City, UT, [4]Texas Children's Hosp., Houston, TX, [5]Univ. of Brasilia, Brasilia, Brazil, [6]Robinow Syndrome Fndn., Anoka, MN, [7]Human Genome Sequencing Ctr., Houston, TX, [8]GeneDx Inc., Gaithersburg, MD, [9]UTHlth., Houston, TX, [10]Oxford Univ. Hosp. NHS Fndn. Trust, Oxford, United Kingdom, [11]NHS Lothian, Edinburgh, United Kingdom, [12]Botnar Res. Ctr., Oxford, United Kingdom, [13]Great Ormond Street Hosp., London, United Kingdom, [14]Univ. Clinical Ctr. Zagreb, Zagreb, Croatia, [15]Univ. of Manchester, Sale, United Kingdom, [16]Manchester center for genomic medicine, Manchester, United Kingdom, [17]Baylor Col Med/TX Child Hosp, Houston, TX, [18]Columbia Univ., New York, NY, [19]Arnold Palmer Hosp. for Children, Orlando, FL, [20]Duke Univ. Med. Ctr., Durham, NC, [21]Inova Translational Med. Inst., Falls Church, VA, [22]Cook Children's Hosp., Fort Worth, TX, [23]UNC - Chapel Hill, Chapel Hill, NC, [24]State Univ. of Rio de Janeiro, Rio de Janeiro, Brazil, [25]Inst. da CrianÃ§a, SÃ£o Paulo, Brazil, [26]Baylor Coll. Med., Houston, TX, [27]Baylor Genetics Lab., Houston, TX, [28]Pacific Northwest Res. Inst., Seattle, WA

Disclosure Block: **C. Zhang:** None.

Robinow syndrome (RS) is a genetically heterogeneous disorder with six genes, *DVL1, DVL3, FZD2, NXN, ROR2,* and *WNT5A*, shown to be involved to date; the gene products converge on the WNT/planar cell polarity (PCP) signaling pathway. RS is a skeletal dysplasia with a recognizable pattern of craniofacial and physical characteristics. To further explore the genetic heterogeneity, paralog contribution, and phenotypic variability of RS, we investigated a cohort of 22 individuals clinically diagnosed with RS from 18 unrelated families. Pathogenic or likely pathogenic variants in genes associated with RS, or Mendelian disorders with phenotypic overlap with RS, were identified in all 22 individuals; notably, one individual was found to have a variant in *DVL2,* which has not previously been associated with RS or any other phenotype. We retrospectively collected medical records of subjects in a subgroup of this cohort (N = 16) and examined clinical descriptions of published cases (N = 52) and used these data to perform HPO term-based quantitative phenotypic analyses to investigate paralog contributions and allele-specific phenotypic differences. Individuals with *FZD2* variants clustered into two groups that correlated with pathogenic allele type (missense or truncating), in general individuals having missense variants in *FZD2* are overall more severely affected than those with truncating variants. Moreover, probands with biallelic *NXN* variants clustered together with the majority of probands carrying *DVL1, DVL2,* and *DVL3* variants, blurring the phenotypic distinction between the *NXN*-autosomal recessive and dominant forms of RS. In addition, disease-to-gene and disease-to-disease HPO analyses revealed that RS can be quantitatively distinguished from diseases caused by variants affecting non-RS-associated paralogs of *WNTs*, *FZDs*, or *RORs*; supporting the hypothesis that RS results from perturbation of a very specific pathway. Our study: i) identifies previously undescribed pathogenic variant alleles, ii) provides evidence for a new RS â□□disease geneâ□□ *DVL2*, iii) examines paralog specificity of trait manifestations, iv) delineates allele-specific phenotypic differences, v) provides biological insights into a genetically heterogenous

disorder, and vi) reveals the power of deep phenotypic analyses to tease apart genetically heterogeneous disorders.

PrgmNr 3409 - *De novo TNPO2* variants are associated with developmental delays, neurological issues and dysmorphic features in humans and alter *TNPO2* activity in *Drosophila*

View session detail

Author Block: **L. Goodman**[1], H. Cope[2], Z. Nil[1], T. A. Ravenscroft[1], W-L. Charng[3], S. Lu[1], A-C. Tien[1], R. Pfundt[4], D. A. Koolen[5], C. A. Haaxma[6], H. E. Veenstra-Knol[7], J. S. Klein Wassink-Ruiter[7], M. R. Wevers[8], M. Jones[9], L. E. Walsh[10], V. H. Klee[11], M. Theunis[12], E. H. Legius[13], D. Steel[14], K. E. S. Barwick[14], M. A. Kurian[15], S. S. Mohammad[16], R. C. Dale[17], P. A. Terhal[18], E. van Binsbergen[18], B. Kirmse[19], B. Robinette[19], B. CognÃ©[20], B. Isidor[21], T. A. Grebe[22], P. Kulch[23], B. E. Hainline[24], K. Sapp[25], E. Morava[26], E. W. Klee[27], E. Macke[27], P. S. Trapane[28], C. Spencer[29], Y. Si[30], A. H. Begtrup[31], M. Moulton[32], D. Dutta[32], O. Kanca[32], M. F. Wangler[33], S. Yamamoto[34], H. J. Bellen[32], K. Tan[35]; [1]Neurological Res. Inst., Houston, TX, [2]Duke Univ. Med. Ctr., Durham, NC, [3]Washington Univ. in St. Louis, St. Louis, MO, [4]Radboud Univ. Nijmegen Med. Ctr., Nijmegen, Gelderland, Netherlands, [5]Radboud Univ. Med. Ctr., Nijmegen, Gelderland, Netherlands, [6]Amalia Childrenâ□□s Hosp., Nijmegen, Netherlands, [7]Univ. Med. Ctr. Groningen, Groningen, Netherlands, [8]Radboud Univ. Med. Ctr., Nijmegen, Netherlands, [9]Houston Area Pediatric Neurology, Houston, TX, [10]Indiana Univ Sch Med., Indianapolis, IN, [11]Riley Hosp. for Children, Indianapolis, IN, [12]Univ. Hosp. Leuven, Leuven, Belgium, [13]Univ. Hosp. Leuven, LEUVEN, Belgium, [14]UCL Great Ormond Street Inst. of Child Hlth., London, United Kingdom, [15]London, United Kingdom, [16]The Children's Hosp. at Westmead, Westmead, Australia, [17]Sydney Med. Sch., Westmead, Australia, [18]Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, [19]Univ. of Mississippi Med. Ctr., Jackson, MS, [20]CHU NANTES, NANTES, France, [21]CHU Nantes, Nantes, France, [22]Phoenix Children's Hosp., Phoenix, AZ, [23]Phoenix Childrenâ□□s Hosp., Phoenix, AZ, [24]Indiana Sch Med, Indianapolis, IN, [25]Indiana Univ. Sch. of Med., Indianapolis, IN, [26]Mayo Clinic Rochester, Rochester, MN, [27]Mayo Clinic, Rochester, MN, [28]UF Coll. of Med. - Jacksonville, Jacksonville, FL, [29]Univ. of Florida Coll. of Med., Jacksonville, FL, [30]GeneDx, Gaithersburg, MD, [31]GeneDx, Inc., Cincinnati, OH, [32]Baylor Coll. of Med., Houston, TX, [33]Baylor Coll. Med., Houston, TX, [34]Baylor Coll. of Med., HOUSTON, TX, [35]Duke Univ. Med. Ctr., Div. of Med. Genetics, Durham, NC

Disclosure Block: **L. Goodman:** None.

Transportin-2 (TNPO2) mediates multiple pathways including non-classical nucleocytoplasmic shuttling of >60 cargoes, including developmental and neuronal proteins. We identified fifteen individuals carrying *de novo* coding variants in *TNPO2* who presented with global developmental delay (GDD), dysmorphic features, ophthalmologic abnormalities, and neurological features. To assess the nature of these variants, functional studies were performed in *Drosophila*. We found that fly *dTnpo* (orthologous to *TNPO2*) is expressed in a subset of neurons. *dTnpo* is critical for neuronal maintenance and function as downregulating *dTnpo* in mature neurons using RNAi disrupts neuronal activity and survival. Altering the activity and expression of *dTnpo* using mutant alleles or RNAi causes developmental defects, including eye and wing deformities and lethality. These effects are dosage-dependent as more severe phenotypes are associated with stronger *dTnpo* loss. Interestingly, similar phenotypes are observed with *dTnpo* upregulation and ectopic expression of TNPO2, showing that loss and gain of Transportin activity causes developmental defects. Further, proband-associated variants can cause more or less severe developmental abnormalities compared to wild-type TNPO2 when ectopically expressed. The impact of the variants tested seems to correlate with their position within the protein. Specifically, those that fall within the RAN binding domain cause more severe toxicity and those in the acidic loop are less toxic. Variants within the cargo binding domain show tissue-dependent effects. In summary, *dTnpo* is an essential gene in flies during development and in neurons. Further, proband-associated *de novo* variants within *TNPO2* disrupt the function of the encoded protein. Hence, TNPO2-variants are causative for neurodevelopmental abnormalities.

PrgmNr 3410 - *De novo* damaging variants in the microRNA processor *DROSHA* are associated with a severe progressive neurological disorder

View session detail

Author Block: **S. Barish**[1], M. Senturk[1], K. Schoch[2], A. Minogue[3], J. Seeman[3], Undiagnosed Diseases Network, M. F. Wangler[4], S. Arur[3], Y-h. Jiang[5], S. Yamamoto[1], V. Shashi[6], H. J. Bellen[7]; [1]Baylor Coll. of Med., HOUSTON, TX, [2]Duke Hlth., Durham, NC, [3]Univ. of Texas MD Anderson Cancer Ctr., HOUSTON, TX, [4]Baylor Coll. Med., Houston, TX, [5]New Haven, CT, [6]Duke Univ Med Ctr, Durham, NC, [7]Baylor Coll. of Med., Houston, TX

Disclosure Block: **S. Barish:** None.

DROSHA encodes a ribonuclease that is a subunit of the Microprocessor complex and is involved in the first step of microRNA (miRNA) biogenesis. To date, *DROSHA* has not yet been associated with a Mendelian disease. Here we describe two individuals with profound intellectual disability, epilepsy, white matter atrophy, microcephaly, and dysmorphic features, who carry damaging *de novo* heterozygous variants in *DROSHA*. *DROSHA* is constrained for missense variants and moderately intolerant to loss of function (o/e=0.24). The loss of the fruit fly ortholog *drosha* causes developmental arrest and death in third instar larvae, a severe reduction in brain size, and loss of imaginal discs in the larva. Loss of *drosha* in eye clones causes small and rough eyes in adult flies. One of the identified *DROSHA* variants (p.D1219G) behaves as a strong loss-of-function allele in flies, while another variant (p.R1342W) is less damaging in our assays. In worms, a knock-in that mimics the p.D1219G variant at worm equivalent residue causes loss of miRNA expression and heterochronicity, a phenotype characteristic of the loss of miRNA. Together, our data show that the *DROSHA* variants found in the individuals presented here are damaging based on functional studies in model organisms and loss of this gene in human and fly leads to a severe phenotype involving the nervous system.

PrgmNr 3411 - *De novo* variant in *MRTF-B* is associated with intellectual disability, minor dysmorphic features, expressive language delay, impulse control issues, and fine motor delay

View session detail

Author Block: **J. C. Andrews**[1], O. Kanca[1], D. Li-Kroeger[1], C. Tifft[2], E. Macnamara[3], S. Yamamoto[1], H. J. Bellen[1], M. C. Malicdan[4], M. F. Wangler[1]; [1]Baylor Coll. of Med., Houston, TX, [2]NIH, Bethesda, MD, [3]NIH/NHGRI, Arlington, VA, [4]NIH/NHGRI, Rockville, MD

Disclosure Block: **J.C. Andrews:** None.

Myocardin-related transcription factor B (MRTF-B) is a member of a family of genes which serve to potentiate serum response factor (SRF)-dependent transcription and is highly conserved in both vertebrate and invertebrate model organisms. The MRTF-B protein is not currently associated with a human disease but has been shown to be highly expressed in all human tissues save the lung. Here we report a proband with a *de novo* variant in the second RPEL domain of *MRTF-B* with intellectual disability, minor dysmorphic features, expressive language delay, impulse control issues, and fine motor delay. We have generated a fly model of human *MRTF-B*, using a UAS construct carrying either the human reference or variant cDNA. Expression of the UAS-human variant cDNA under the control of the *Nubbin-Gal4* driver was sufficient to induce significant morphological changes in the wing, while the expression of the human control reference cDNA produced only minor changes in the wings posterior crossvein. Expression of *Drosophila Mrtf* using the *Nubbin-Gal4* driver produced a similar change in crossvein length as was observed with the human reference. In *Drosophila*, the SRF ortholog is known as *blistered (bs)* and is known to suppress wing vein formation and promote the development of intervein cells. As exogenous co-expression of *bs* and *Mrtf* has been previously shown to significantly alter wing morphology, we also expressed our human reference and variant cDNA lines concurrently with a UAS-*bs* line. In these experiments we found that wing morphology was highly disrupted when *bs* and the reference human cDNA were co-expressed, while the co-expression of human variant cDNA and *bs* was lethal. To further clarify our findings, we sought and identified two additional probands with variants within or near the second RPEL domain. Expression of these additional variants using *Nubbin*-Gal4 produced changes in the posterior crossvein analogous to the changes observed when our human reference line was expressed. Taken together, these findings suggest that different residue changes within the RPEL domain of *MRTF-B* can have drastically different morphological effects in the fly wing, and our initial *de novo* variant may underly a novel disorder.

PrgmNr 3412 - *SLC6A1 Portal: Genetic variant analysis and educational resource for SLC6A1-related disorders*

View session detail

Author Block: **A. Stefanski**[1], E. PÃ©rez-Palma[2], T. BrÃ¼nger[3], C. KlÃ¶ckner[4], M. Macnee[3], P. May[5], J. R. Lemke[6], K. Goodspeed[7], J-Q. Kang[8], K. Johannesen[9], R. MÃ¸ller[9], A. Freed[10], D. Lal[11,12]; [1]Genomic Med. Inst. & Epilepsy Ctr. Cleveland Clinic, Cleveland, OH, [2]Univ. del Desarrollo, Santiago, Chile, [3]Cologne, Germany, [4]Univ. of Leipzig Med. Ctr., Leipzig, Germany, [5]LCSB, Esch-Belval, Luxembourg, [6]Leipzig, Leipzig, Germany, [7]Children's Hlth., Med. Ctr., UT Southwestern, Dallas, TX, [8]Dept. of Neurology, Vanderbilt Univ. Med. Ctr., Nashville, KY, [9]The Danish Epilepsy Ctr., Dianalund, Denmark, [10]Children's Hosp. of Philadelphia, Philadelphia, PA, [11]Cleveland Clinic, Cleveland, OH, [12]Stanley Ctr. of Psychiatric Res., Broad Inst. of Harvard and MIT, Cambridge, MA
Disclosure Block: **A. Stefanski:** None.

Genetic variants in the *SLC6A1* gene, encoding the GABA transporter protein type 1 (GAT1), are among the most frequent causes of complex neurodevelopmental disorders, featuring epilepsy, autism spectrum disorder, and intellectual disability. Currently, clinical, genetic, and molecular data about the *SLC6A1* gene is not connected and distributed across registries, databases, and the literature. To facilitate further research, the translation of research findings into improved patient management and overall awareness, we developed the *SLC6A1*-Portal: A novel online resource containing a unique expert-curated *SLC6A1* dataset alongside user-friendly semi-automated data visualizations. In addition, we incorporated extensive educational resources for families, clinicians, and analytical applications to translate current scientific knowledge aiming to improve the quality of genetic counseling and research opportunities for scientists.We gathered clinical, genetic, and molecular data from 150 *SLC6A1* disorder patients. We also aggregated data from 22 functionally tested variants. In addition, we produced several short educational videos about *SLC6A1*-related disorders, genetics, precision medicine, and related subjects. These videos were translated in >10 languages to mobilize underreported individuals worldwide and create a comprehensive resource for families, researchers, and clinicians.We designed the *SLC6A1*-Portal for three user scenarios: i) Education: We provide a rich information resource for *SLC6A1*-related disorders for professionals and the general public. ii) Variant interpretation: Access to expert-level variant interpretation through novel expert-curated web applications that enable variant pathogenicity classification and genotype-phenotype correlation. iii) Research resource: Throughout the *SLC6A1*-Portal, we provide novel research tools to explore related biomedical data resources. Users will be guided by tutorials with illustrative examples and will not require any programming skills.The *SLC6A1*-Portal infrastructure is scalable and, with increasing data, will guide variant interpretation, research, and education for *SLC6A1*. The *SLC6A1*-Portal will be hosted at http://slc6a1-portal.broadinstitute.org at the time of the meeting. We are actively looking for collaborators contributing to this project.

PrgmNr 3414 - A novel *de novo* variant in *PPP2CA* affecting the 5â□□UTR region causes a syndrome characterized by global developmental delay with specific structural brain abnormalities, congenital microcephaly and dysmorphic features

View session detail

Author Block: **A. Cassidy**[1], K. Treat[1,2], S. Pelletier[1], E. Conboy[1,2], F. Vetrini[1,2]; [1]Dept. of Med. and Molecular Genetics, Indiana Univ. Sch. of Med., Indianapolis, IN, [2]Undiagnosed Rare Disease Clinic, Indianapolis, IN

Disclosure Block: **A. Cassidy:** None.

Introduction: Heterozygous pathogenic variants in the *PPP2CA* gene cause a recently described neurodevelopmental syndrome characterized by global developmental delay, intellectual disability and highly variable features including hypotonia, seizures, dysmorphic features, behavioral abnormalities, and a wide spectrum of structural brain abnormalities. This is a very rare disorder with only 16 patients described worldwide (Raynhout et al., 2019). *PPP2CA* encodes the PP2A catalytic C-Î± subunit of the large PP2A complex which is highly expressed in the brain and regulates neuronal signaling by catalyzing phospho-serine and threonine de-phosphorylation in a plethora of different substrates. All the patients harbor *de novo* variants in *PPP2CA* predicted to affect the catalytic activity, the binding to other subunits in the complex or the protein levels with both loss-of-function or dominant negative mechanisms.**Methods:** A 2-year-old male presented to Undiagnosed Rare Disease Clinic at Indiana University School of Medicine/IUHeath with polymicrogyria, white matter loss resulting in ventriculomegaly, severe upper airway disease, hypotonia, feeding dysfunction and dysmorphic features including hypertelorism, wide nasal bridge, congenital microcephaly, and deeply set eyes. Whole-Exome Sequencing analysis revealed a *de novo* VUS c.-1_2delCAT in *PPP2CA* gene, affecting the 5â□□UTR and absent from public databases of normal variation. To further investigate the variant, we performed *in silico* predictions, and mRNA and protein studies to assess the level of expression. **Results**:The RT-qPCR performed in mRNA extracted from peripheral blood showed no reduction of PPP2CA mRNA levels. *In silico* analyses predicted that c.-1_2delCAT deletion alters significantly the natural Kozak sequence (GCATCATGG>GCATG). Subsequent western blot analysis performed by transfecting HEK293T cells with WT-PPP2CA-HA and 1_2delCAT-PPP2CA-HA, revealed the presence of an alternative 29 KDa N-terminal truncated product caused by a differential usage of a downstream in frame alternative ATG in position +66. This deleted product does not possess the previously characterized critical residues D57,H59 and G60 and could act as a dominant negative or hypomorphic allele. **Conclusion:** Here we report for the first time a novel variant affecting the 5â□□UTR region of the *PPP2CA* gene in a patient with a PPP2CA-related disorder. Taken together, our findings expand the phenotypic and genotypic spectrum of the recently described PPP2CA-related syndrome and illustrates how the analysis of non-coding variants can help inform the disease-causing mechanism and improve the diagnostic yield.

PrgmNr 3415 - Allelic frequency and distribution of *M98K*variant associated with adult-onset open angle glaucoma in Puerto Rico

View session detail

Author Block: **C. PagÃin**, E. Ramirez Aponte, J. Martinez Cruzado; Univ. of Puerto Rico at Mayaguez, Mayaguez, Puerto Rico
Disclosure Block: **C. PagÃin:** None.

Glaucoma is a severe neurodegenerative, hereditary condition that is increasingly prevailing in native Puerto Rican citizens. The purpose of this study is to correlate the *M98k* (rs11258194) missense variant in the *Optineurin* (*OPTN*) protein with Adult-Onset Primary Open Angle Glaucoma (POAG) in Puerto Rico; in addition, determine the distributional mutated allelic frequency throughout the Island. *Optineurin* is an adaptive protein that interacts in various nervous cell processes such as vesicular transportation, signaling and autophagy. By selecting 625 patient samples that span a total of 30 municipalities as a guide of the general population, we were able to genotype the samples by RT-PCR with a TaqMan Assay for SNP rs11258194. The samples were further analyzed through geographic distribution via MapViewer, a map-rendering tool. The results showed a total of 60 mutated alleles (48 heterozygous, 6 homozygous) mainly distributed along the coastline of Puerto Rico. Therefore, heterozygous patients with *M98k* variant present are more prone than wildtype patients to develop a neurodegenerative condition. Further studying will confirm if homozygous patients present POAG or other nervous-related condition such as Amyotrophic Lateral Sclerosis.

PrgmNr 3416 - An integrated machine learning and functional analysis approach for resolution of variants of uncertain significance (VUS) in *STXBP1*

View session detail

Author Block: **J. Calhoun**[1], J. Gunti[1], H. C. Mefford[2], S. Schnell[3], V. Aguiar-Pulido[4], M. E. Ross[5], J. Parent[3], L. L. Isom[3], EpiMVP Consortium, G. L. Carvill[1]; [1]Northwestern Univ., Chicago, IL, [2]St. Jude Children's Res. Hosp., Memphis, TN, [3]Univ. of Michigan, Ann Arbor, MI, [4]Univ. of Miami, Miami, FL, [5]Weill Cornell Med., New York, NY

Disclosure Block: **J. Calhoun:** None.

Variants of uncertain significance (VUS) pose a significant challenge for genetic diagnosis of epilepsy, in that these variants can be classified as neither pathogenic nor benign. In order to address this challenge, we established a highly integrated Epilepsy Multiplatform Variant Predictor (EpiMVP) Center Without Walls to develop precise, single-gene *in silico* pathogenicity prediction tools (EpiPred) for the most common epilepsy-associated genes. We are using machine learning algorithms to discriminate pathogenic from benign variants. This approach will be bolstered by functional characterization of a subset of variants in multiple cellular and animal models. The first gene for study is *STXBP1*, a gene implicated in a range of pediatric-onset epilepsies. We first curated a robust training set for the development of our classifier. This consisted of known benign or likely benign (BLB) missense *STXBP1* variants from the general population (n=128) and known pathogenic or likely pathogenic (PLP) missense *STXBP1* variants from multiple sources, including industry partners (Invitae, GeneDx, Geisenger), Clinvar, and the literature (n=93). Variants were annotated with multiple features (n=19) using *in silico* measures related to evolutionary conservation, protein structure and stability, among others. Our approach employed standard unsupervised and supervised machine learning algorithms to develop an optimal single-gene classifier. Our classifier outputs a score ranging from PLP (high) to BLB (low) and was used to score *STXBP1* VUSs (n=133) collated from ClinVar and industry partners. Our classifier is successful at discriminating *STXBP1* missense BLB and PLP variants (ROC_{AUC} > 0.9). We used four different models to score and then rank each of the 133 VUS. About half of the VUSs (50.3%) were highly concordant across the four models (SD 0.1). We selected eight representative variants for functional modeling. The first iteration of our EpiPred classifier can distinguish between BLB and PLP variants with high specificity. After functional validation in the EpiMVP cellular and animal models, we will improve the predictive power of EpiPred and release this model to the wider scientific and epilepsy community for interpretation of VUS. Our overarching goal is to develop gene-specific classifiers for additional epilepsy-associated genes with an emphasis on non-ion channel genes with the highest volume of VUS and PLP variants identified in routine clinical genetic testing.

PrgmNr 3417 - Biallelic *NAV2* truncating variants cause a neurodevelopmental disorder with cerebellar cortical dysplasia

View session detail

Author Block: **J. A. Rosenfeld**[1], A. Accogli[2,3], I. Musante[3,4], P. Scudieri[3,4], M. Severino[3], S. Baldassari[3], M. Iacomino[3], A. Riva[4], G. Balagura[4], G. Piccolo[4,3], C. Minetti[4,3], D. Roberto[5], F. Xia[6,7], R. Razak[8], E. Lawrence[8], M. Hussein[8], E-H. Chang[9], M. Holick[1,8], Undiagnosed Diseases Network, L. T. Emrick[6,8], P. J. A. McCaffery[10], M. Clagett-Dame[11], S. R. Lalani[6], F. Zara[12,4], P. Striano[4,3], V. Salpietro[4,3,13]; [1]Baylor Coll. of Med., Houston, TX, [2]McGill Univ., Montreal, QC, Canada, [3]IRCCS Istituto Giannina Gaslini, Genoa, Italy, [4]Univ. of Genoa, Genoa, Italy, [5]Tor Vergata Univ. of Rome, Rome, Italy, [6]Baylor Coll. Med., Houston, TX, [7]Baylor Genetics Lab., Houston, TX, [8]Texas Children's Hosp., Houston, TX, [9]Retina and Vitreous of Texas, Houston, TX, [10]Univ. of Aberdeen, Aberdeen, United Kingdom, [11]Univ. of Wisconsin, Madison, WI, [12]Inst G Gaslini, Genova, Italy, [13]Univ. Coll. London, London, United Kingdom
Disclosure Block: **J.A. Rosenfeld:** Other; Baylor Genetics.

Cerebellar hypoplasia and dysplasia encompass a group of clinically and genetically heterogeneous disorders frequently associated with neurodevelopmental impairment. We report a 7-year-old female with global developmental delay (mostly mild, with gross motor skills most significantly impacted), abnormal brain MRI (marked cerebellar vermis hypoplasia, bilateral cerebellar foliation defects, pontine hypo-dysplasia, splayed thin superior cerebellar peduncles with a molar tooth-like configuration, corpus callosum hypodysgenesis, absent anterior commissure, diffuse dysgyria, agenesis of the olfactory bulbs, mild optic nerve hypoplasia, and enlarged dysmorphic lateral ventricles), ophthalmologic abnormalities (atypical chorioretinal scarring, retinal ischemia, retinal neovascularization, and retinal detachment), polyvalvular heart disease, and dysmorphic features. Trio-based clinical exome sequencing revealed compound heterozygous truncating variants in the neuron navigator 2 (*NAV2*) gene, which was not previously associated with human disease. *NAV2* encodes a member of the neuron navigator protein family, widely expressed within the central nervous system (CNS), particularly abundant in the developing cerebellum, with pivotal functions in cytoskeletal dynamics and neurite outgrowth. *Nav2* hypomorphic mice display cerebellar hypoplasia with abnormal foliation, due to impaired axonal outgrowth. To further explore the possible connection of *NAV2* to human disease, we performed expression studies and migration assays in patient-derived fibroblasts, which confirmed absence of full-length NAV2 and showed perturbed cellular migration. We also evaluated the overall CNS histopathology of the *Nav2* hypomorphic mouse, which revealed multiple developmental anomalies (i.e., cerebellar hypoplasia and dysplasia and corpus callosum abnormalities), strikingly overlapping our probandâ□□s brain findings. Together, these data support recessive loss of function of *NAV2* as a cause for a neurodevelopmental disorder with complex brain malformations, particularly in the cerebellum.

PrgmNr 3418 - Delineation of a neurodevelopmental syndrome caused by *PAX5* haploinsufficiency

View session detail

Author Block: **Y. Gofin**[1], T. M. Scott[2], C. A. Genetti[3], P. B. Agrawal[4], J. D. Picker[5], M. H. Wojcik[4,6], S. Lynch[7], S. W. Scherer[8,9], J. Howe[8], M. A. Gillentine[10,11], T. Wang[12], E. E. Eichler[10,13], R. K. Earl[12], S. DiTroia[6], W. D. Graf[14], D. A. Scott[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Brigham Young Univ., Provo, UT, [3]Boston Children's Hosp, Boston, MA, [4]Boston Children's Hosp. and Harvard Med. Sch., Boston, MA, [5]Children's Hosp, Boston, MA, [6]Broad Inst. of MIT and Harvard, Cambridge, MA, [7]Children's Hlth.Ireland, Dublin, Ireland, [8]The Hosp. for Sick Children, Toronto, ON, Canada, [9]Univ. of Toronto, Toronto, ON, Canada, [10]Univ. of Washington Sch. of Med., Seattle, WA, [11]Seattle Childrenâ□□s Hosp., Seattle, WA, [12]Univ. of Washington, Seattle, WA, [13]Howard Hughes Med. Inst., Univ. of Washington, Seattle, WA, [14]Connecticut Childrenâ□□s, Univ. of Connecticut, Hartford, CT

Disclosure Block: **Y. Gofin:** None.

The human PAX (Paired Box) family of transcription factors has nine members, all of which are associated with a genetic syndrome, susceptibility to a disease, or a structural birth defect. Currently, the only known disease association for the gene *PAX5* is susceptibility to acute lymphoid leukemia. However, *PAX5* is transiently expressed during embryogenesis in the central nervous system, is highly loss-of-function intolerant (pLI = 1, o/e = 0) and missense constrained (Z = 2.54) in gnomAD, is associated with abnormal posterior midbrain and cerebellum development in mice, and has been suggested as a candidate gene for autism and intellectual disability based on data from large cohort studies. In this study, we describe eight individuals who carry putatively deleterious *PAX5* variants. These include two individuals with interstitial deletions that include *PAX5*, two individuals with frameshift variants that are likely to trigger nonsense-mediated mRNA decay, and four individuals with single nucleotide variants, three of which affect the *PAX5* DNA binding domain. All the individuals in this cohort had developmental delay, intellectual disability and/or autism spectrum disorder. Speech development was particularly affected, with many showing regression in their expressive language abilities and approximately half being non-verbal. Feeding issues and behaviorally-defined psychiatric diagnoses were common, and two individuals have seizures. No structural abnormalities were noted in those who had brain imaging, and no clear pattern of dysmorphic features was apparent. Our findings suggest that *PAX5* haploinsufficiency causes a neurodevelopmental disorder characterized by severe deficits in expressive language development, variable intellectual disability, and autism spectrum disorder.

PrgmNr 3419 - Diagnostic rate and clinical implications of exome sequencing in adults with intellectual disability

View session detail

Author Block: **A. Sabo**[1], D. R. Murdock[2], S. Pereira[3], S. Dugan[1], M. Gingras[3], D. M. Muzny[3], R. A. Gibbs[3]; [1]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. of Med., Houston, TX, [3]Baylor Coll. Med., Houston, TX

Disclosure Block: **A. Sabo:** None.

Establishing a genetic diagnosis for individuals with intellectual disability (ID) benefits patients and their families as it may inform the prognosis, lead to appropriate therapy, and facilitate access to medical and supportive services. Exome sequencing has been successfully applied in a diagnostic setting, but most clinical exome referrals are pediatric patients, with many adults with ID lacking a comprehensive genetic evaluation. Many adults living with ID might now be able to be diagnosed with modern methods but are not readily accessing the tests due to a lack of access to care, insurance coverage or knowledge of new opportunities. Our aim is to reach this underserved population by recruiting 100 adults with ID and their families. Recruitment sites include non-clinical organizations that provide education and support services for individuals with ID, as well as three clinical sites: Childrenâ□□s Hospital of San Antonio, Emory University, and Transition Medicine Clinic at Baylor College of Medicine. At all sites adults with ID of unexplained etiology with all levels of severity are included to evaluate a low barrier, broad application clinical genetic testing utilizing WES methods. We utilized exome sequencing and analysis, identified SNVs, Indels, and CNVs and performed clinical variant interpretation for each recruited family. Families where exome sequencing analysis identified pathogenic or likely pathogenic variants were invited for an appointment with a board-certified clinical geneticist to discuss the results and any potential medical follow-up. In addition to establishing diagnostic rates we will formally evaluate the impact of a genetic diagnosis on patients and their families, and the clinical utility, including measures related to diagnosis, patient clinical management, and familial and psychosocial implications using recently developed C-GUIDE.

Our published data from the pilot included 5 families of which 3 received a genetic diagnosis. Currently, we have recruited 15 additional families. For one family we have identified a pathogenic stop-gain variant in the POGZ gene. The individual has ID, autism, and visual and hearing impairment, fitting the diagnosis of White-Sutton syndrome. The samples from remaining 14 families are currently undergoing sequencing, and findings from this next phase will be described in this presentation. Our study demonstrates that there is interest in families with adults with ID in receiving a genetic diagnosis and suggests a high yield of exome sequencing (4 out of 6 families) as a diagnostic tool in adult patients with ID who have not undergone comprehensive sequencing based genetic testing.

PrgmNr 3420 - Expanding the Genotypic and Phenotypic Spectrum of CTCF-Related Disorders: 19 new patients andreviews of published literature

View session detail

Author Block: **H. Valverde de Morales**, P. Piwko, J. DiMonda, H-L. Wang, V. Corces, H. Li; Emory Univ. Sch. of Med., Atlanta, GA
Disclosure Block: **H. Valverde de Morales:** None.
Background Monoallelic variants in *CTCF* cause an autosomal dominant neurodevelopmental disorder with a wide range of symptoms affecting brain, craniofacial features, growth and various anomalies. Given the increased application of exome sequencing, a continuously growing number of patients has been identified and further delineation of the clinical spectrum of CTCF-related disorders (CTCFRD) is needed. **Methods** Individuals with monoallelic variants in *CTCF* were recruited through a CTCF family support group, Genematcher and a commercial genetic diagnostic laboratory. The detailed clinical information along with patientâ□□s facial photos at various ages were collected through RedCap under Emory IRB approval. Clinical variables were analyzed across our novel patients and previously published patients. Facial profiles were investigated via DeepGestalt using the Face2Gene tool. **Results** In our study, we have enrolled 31 patients, clinical information was completed for 20, including 19 novel patients from 17 unrelated families and 1 previously reported patient who has updated information. We identified 12 novel variants (4 frameshift, 1 nonsense, 5 missense and 2 small in-frame deletions), which expanded the total variants in *CTCF* from 45 up to 57. Though most variants are de novo, we report one familial case affecting three generations. Common clinical features in our cohort include developmental delay or intellectual disability (18), failure to thrive or feeding difficulties (12), sleeping difficulties (11), intrauterine growth restriction/small gestational age (IUGR/SGA) (11), skeletal anomalies (10), genitourinary anomalies (9), and hypotonia (9). Noticeably, the incidence of autism (9) was higher in our cohort compared to the existing literature (45% vs 30%). Other less common anomalies include teeth abnormalities (7), seizures (5) and palatal abnormalities (4). Some features, such as cardiac defects (4) and microcephaly (0), are not as common as previously reported. Dysmorphology analysis of patients identified broad nasal bridge, deep-set eyes, cupped ear and pointed chin as the hallmark features of CTCFRD. DeepGestalt training from 36 patients at various ages (18 patients from our study and 18 from the literature) is still ongoing. **Conclusions** Our study expands the genotypic and phenotypic spectrum of CTCFRD that can guide genetic counseling and surveillance care. Both dysmorphology and digital analysis could facilitate the early diagnosis of CTCFRD.

PrgmNr 3421 - Expanding the phenotypic spectrum of *EXOSC8*-related neurological disease: Presentation of an Egyptian family with neurodevelopmental disorder due to a novel *EXOSC8* missense variant

View session detail

Author Block: **I. Herman**[1,2,3], D. G. Calame[1,2,3], R. DUAN[2], T. Mitani[2], J. M. Fatih[2], A. K. Saad[2,4], A. van Hoof[5], J. E. Posey[2], D. Pehlivan[1,2,3], G. M. H. Abdel-Salam[6], D. Marafi[2,7], J. R. Lupski[2,3,8,9]; [1]Div. of Neurology and Dev.al NeuroSci., Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, [2]Molecular & Human Genetics, Baylor Coll. of Med., Houston, TX, [3]Texas Childrens Hosp., Houston, TX, [4]4.Med. Molecular Genetics Dept., Natl. Res. Ctr., Cairo, Egypt, [5]5.Dept. of Microbiol. and Molecular Genetics, McGovern Med. Sch. at The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [6]Clinical Genetics Dept., Human Genetics and Genome Res. Div., Natl. Res. Ctr., Cairo, Egypt, [7]Dept. of Pediatrics, Faculty of Med., Kuwait Univ., Safat, Kuwait, [8]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [9]Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX

Disclosure Block: **I. Herman:** None.

Background: The exosome is a protein complex essential for RNA processing and degradation with a critical role in gene expression. To date, 6 *EXOSC* genes (*EXOSC1*, *EXOSC2*, *EXOSC3*, *EXOSC5*, *EXOSC8*, *EXOSC9*) are linked to human disease. Variants in *EXOSC8* are associated with autosomal recessive (AR) pontocerebellar hypoplasia 1C (MIM #616081), described in only 3 families, which is characterized by a universally progressive infantile-onset neurodegenerative disease with muscle atrophy, brain abnormalities, and death in infancy. Here, we expand the phenotypic spectrum of *EXOSC8*-related disease and describe the clinical and molecular features of a 4th family with neurodevelopmental disorder and survival into adulthood in association with a seemingly milder missense allele.

Materials and Methods: The family was recruited as part of the Baylor-Hopkins Center for Mendelian Genomics research initiative. Study consent was obtained from all participants. Trio exome sequencing (ES) with family-based rare variant analysis was performed.

Results: The proband is a 5 year-old girl born to first cousin Egyptian parents. Early clinical features included neonatal jaundice, respiratory distress, aspiration, cyanosis, and global developmental delay. Neurological examination showed dysarthria, hyperreflexia, dysmetria, intention tremor, borderline intellect, and no signs of amyotrophy or neuropathy. Brain MRI showed cerebellar vermis atrophy. She has an adult maternal aunt and uncle with similar phenotypic features with intellectual disability and wheelchair-bound status. Trio ES revealed a novel homozygous *EXOSC8* missense variant (c.398G>C;p.Trp133Ser), predicted damaging by multiple *in silico* analyses, and the affected amino acid residue is fully conserved across species. Further analysis revealed evidence for identity-by-descent (IBD) with a total absence of heterozygosity (AOH) of 280.1 Mb and AOH block of 16 Mb surrounding the detected variant. Sanger sequencing and segregation analyses confirmed the homozygous status in the proband and affected aunt and uncle.

Conclusion: The exosome complex is essential in human development as pathogenic variations in several exosome cause human disease. Here we provide evidence that â□□milderâ□□ pathogenic *EXOSC8* variant alleles cause neurodevelopmental disabilities with survival into adulthood. As *EXOSC*-related disease is extremely rare and few families have been described for each gene, it is essential to identify and longitudinally follow additional affected individuals within a human allelic series to capture and characterize the full genotype-phenotype spectrum of *EXOSC*-related disease.

PrgmNr 3422 - Genetic heterogeneity in *SARS2* gene in a patient with global developmental delay and spastic paraplegia

View session detail

Author Block: **E. Glenn**[1], E. Vasconcellos[2], P. Jayakar[3]; [1]Nicklaus Children's Hosp., Miami, FL, [2]Wemind Inst., Miami, FL, [3]Miami Children's Hosp, Miami, FL
Disclosure Block: **E. Glenn:** None.
Mitochondrial aminoacyl-tRNA synthetases are an increasingly studied family of genes in which mutations may cause phenotypes ranging from isolated tissue disease to distinct syndromes. Mitochondrial seryl-tRNA synthetase gene (*SARS2*) missense mutations have been particularly identified in HUPRA syndrome. Patient is a 2-year-old hispanic male of first cousins consanguineous union, presenting with global developmental delay, short stature, distal hypertonia, poor weight gain and central hypotonia. Birth history remarkable only for IUGR. Developmental delay noted at 7 months of age. At 2 years of age, he cannot sit or roll, is just starting to reach for objects and lifts head while prone; babbles but is non-verbal. Has not had any regressions. Patient presents short stature, plagiocephaly, strabismus, no facial dysmorphism. Neurological exam shows head lag, with a picture of spastic quadriplegia with opisthotonic posture, bilateral cortical thumbs, tight heel cords, hyperreflexia, including cross adductors, ankle clonus and Babinski signs. MRI Brain showed symmetric signal abnormality in the thalami, periventricular white matter with extension to the perirolandic regions. ECHO, EKG, Renal US, blood gas, uric acid, metabolic panel and karyotype were normal. Swallow study showed aspiration in all consistencies. Microarray significant for short runs of homozygosity totaling 19.8 Mb (6.7% of autosomal genome). Exome sequencing positive for *SARS2* homozygous variant c.1347 G>A p.T449, Likely Pathogenic, biparental inheritance; and *PLEKHG2* c.158 C>T p.T531, homozygous, Variant of uncertain significance (VUS), Biparental Inheritance. *SARS2* gene mutation has been associated with HUPRAS, syndrome (hyperuricemia, pulmonary hypertension, renal failure, and alkalosis syndrome) caused by a homozygous mutation in the *SARS2* gene (612804), which encodes mitochondrial seryl-tRNA synthetase, on chromosome 19q13. Linnankivi, T, et al reported a case of homozygous splicing mutation in *SARS2* presenting with progressive spastic paresis and no HUPRAS phenotype. Our patient fits into a clinical phenotype of progressive spastic paresis rather than HUPRAS since he has no hyperuricemia, renal failure or alkalosis with normal cardiac and renal function tests. This demonstrates the genetic heterogeneity of *SARS2* gene and further exemplifies the sensitivity of the nervous system to partially reduced aminoacylation.

PrgmNr 3423 - Genomics of early onset CNS autoimmune demyelinating diseases: Role for deleterious, rare variants in common immune pathways

View session detail

Author Block: **D. G. Calame**[1,2,3], I. Herman[1,2,3], J. M. Fatih[3], K. S. Fisher[1,2], Z. Coban-Akdemir[3,4], H. Du[3], S. N. Jhangiani[5], R. A. Gibbs[3,5], D. Pehlivan[1,2,3], N. Shukla[1], T. E. Lotze[1], J. E. Posey[3], I. K. Chinn[6,7], J. R. Lupski[2,3,5,6]; [1]Div. of Neurology and Dev.al NeuroSci., Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, [2]Texas Childrenâ□□s Hosp., Houston, TX, [3]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [4]Human Genetics Ctr., Dept. of Epidemiology, Human Genetics, and Environmental Sci., Sch. of Publ. Hlth., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [5]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [6]Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX, [7]Section of Immunology, Allergy, and Rheumatology, Texas Childrenâ□□s Hosp., Houston, TX

Disclosure Block: **D.G. Calame:** None.

Background: Demyelinating diseases of the central nervous system (CNS) are poorly understood autoimmune syndromes and include multiple sclerosis (MS), neuromyelitis optica (NMO), acute disseminated encephalomyelitis (ADEM), and chronic lymphocytic inflammation with pontine perivascular enhancement responsive to steroids (CLIPPERS). As in other autoimmune disorders, their genetic underpinnings are felt to be complex, with individual risk reflecting additive or synergistic effects of combinations of common alleles with small effect. In contrast, investigation of the role of rare alleles has been limited, despite their often greater effect on phenotype. Relatedly, rare alleles cause monogenic systemic lupus erythematosus (SLE) and are enriched in patients with early onset disease. With this in mind, we experimentally investigated the hypothesis that rare alleles in immune system genes underlie early onset demyelinating disease.

Methods: We recruited 10 families with early onset (â□¤10 years) demyelinating disease who underwent family-based exome sequencing (ES). ES data were analyzed using rare variant parsing and filtering approach. Family members with systemic autoimmune diseases were recruited for segregation analysis.

Results: Five patients had potentially causative variants in known immune disease genes: *TLR7* (c.1521T>G, p.Phe507Leu, NMO), *TNFAIP3* (c.1069C>T, p.Gln357*, NMO), *IFIH1* (c.2525A>G, p.Glu842Gly, ADEM), *PRF1* (c.657C>A, p.Tyr219*; c.272C>T, p.Ala91Val, CLIPPERS), and *UNC13D* (c.2045G>T, p.Arg682Leu; c.2354T>C, p.Val785Ala [*in cis*], CLIPPERS). Novel disease gene candidates were identified in three patients: *DOCK11* (MS), *SYTL2* (CLIPPERS), and *IFIT1B* (vasculitis). *TNFAIP3*:c.1069C>T segregated with the disease trait within a large autoimmune family with considerable intrafamilial phenotypic variability (median onset: 18 years; range 2-46; diseases: NMO, SLE, Stillâ□□s disease, scleroderma). *TLR7*:c.1521T>G was similarly inherited from a mother with SLE. Immune pathways implicated include type 1 interferon (*TLR7*, *TNFAIP3*, *IFIH1*, *IFIT1B*), cytotoxic lymphocyte degranulation (*UNC13D*, *PRF1*, *SYTL2*), B cell homeostasis (*DOCK11*), and NF-ÎºB (*TLR7*, *TNFAIP3*).

Conclusions: We provide evidence that rare alleles in immune genes underlie early onset demyelinating disorders suggesting â□□medical actionability,â□□ e.g. JAK inhibition for *TNFAIP3* haploinsufficiency. Additional studies in larger cohorts will better delineate the role of rare alleles in demyelinating disease and may uncover the pathophysiological basis of these conditions and additional opportunities for pharmacologic intervention.

PrgmNr 3424 - Identifying variant specific phenotypes for *SLC6A1* related disorders

View session detail

Author Block: **K. Jay**[1], J. C. Andrews[1], S. Jangam[1], H. Pan[1], H. J. Bellen[1], S. Yamamoto[2], M. Wangler[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. of Med., HOUSTON, TX
Disclosure Block: **K. Jay:** None.
The Solute Carrier Family 6, Member 1 (*SLC6A1*) gene was first associated with Doose syndrome in 2015. The patient was carrying a 3p microdeletion and presented with epileptic encephalopathy, intellectual disability (ID), and myoclonic atonic seizures. The first clinical characterization of 34 patients with *SLC6A1* variants was reported in 2018 and the largest case study of patients was presented in 2020. A total of 116 patients with 85 variants were characterized and common phenotypic characteristics include developmental delay, epilepsy, autism spectrum disorder (ASD), and motor dysfunction. Current therapy is focused on seizure suppression however, the GABA transporter type 1 (GAT1) protein has the potential to be targeted directly. Toward this goal, genotype phenotype correlative studies will elucidate the mechanism of transporter dysregulation. We hypothesize both loss of function (LOF) and gain of function mechanisms are responsible for the variable expressivity of *SLC6A1* patient phenotypes. We study *SLC6A1* functional models in *Drosophila melanogaster* utilizing the GAL4-UAS system to express human reference and patient specific variants. We have characterized LOF phenotypes for the fly homolog by evaluating flies with a Trojan GAL4 (TG4) construct, inserted into the first coding intron of the *Drosophila Gat* gene, resulting in severe premature truncation of the *Drosophila* protein. We observe that Gat mutant flies exhibit motor and bang sensitive phenotypes analogous to seizure activity in humans and have a reduced lifespan in homozygous TG4 animals. I will further investigate LOF mechanisms with multiple RNAi targeting *Gat*, utilizing the GAL80ts to delay induction of RNAi in adult flies, allowing us to separate the function of Gat protein throughout development from adult stages. To evaluate patient specific symptoms, we are designing reference and variant constructs for all known recurrent variants previously reported in patients with *SLC6A1* related phenotypes and 5 novel variants collected at the Baylor Genetics lab. Human reference and variant proteins are being expressed with ubiquitous and tissue specific GAL4 driver lines. In parallel we recruit human subjects with *SLC6A1* variants for psychological studies to aid in genotype-phenotype correlation for our functional assays. The models generated will support future studies to predict patient response to GAT1 modulatory therapeutics.

PrgmNr 3425 - KBG Syndrome: Prospective Videoconferencing and Use of AI Facial Recognition in 25 new patients

View session detail

Author Block: **G. LYON**[1], L. Guo[1], J. Park[1], E. Yi[1], T-C. Hsieh[2], P. M. Krawitz[2], E. Marchi[1]; [1]Inst. for Basic Res. in Dev.al Disabilities, Staten Island, NY, [2]Inst. for Genomic Statistics and Bioinformatics, Univ. Bonn, Bonn, Germany

Disclosure Block: **G. Lyon:** None.

Genetic variants in the gene Ankyrin Repeat Domain 11 (ANKRD11) and deletions in 16q24.3 are known to cause KBG Syndrome, a rare syndrome associated with craniofacial, intellectual, and neurobehavioral anomalies. We report 25 unpublished individuals from 22 families, all of whom have a molecularly confirmed diagnosis of KBG syndrome. 21 individuals have de novo mutations, 3 have inherited mutations, and 1 is inherited from a mother exhibiting a low-level of mosaicism. Of these mutations, 20 are truncating (frameshift or nonsense), and 5 are missense. We created a novel protocol for the collection and reporting of the data, including prospectively interviewing these individuals and their families from 8 different countries via videoconferencing by a single clinician, to elucidate the most common phenotypic malformations and developmental anomalies. In addition, the participantsâ medical records, including imaging, were reviewed, and data was compiled and uploaded onto the Human Disease Gene website, with the observed KBG anomalies denoted by Human Phenotype Ontology (HPO) terms. Furthermore, facial photos of the KBG syndrome research participants were obtained and submitted to the Face2Gene website and the GestaltMatcher website. Face2Gene uses convolutional neural networks to identify genetic disorders from a photo by training on thousands of patient images. GestaltMatcher, as an extension of Face2Gene, can quantify the similarity between two patients, enabling us to identify genetic traits and associate them with the most likely genetic disorder. Use of this software may allow for accurate identification of KBG Syndrome using pictures depicting unique craniofacial features. Within our cohort, common phenotypic malformations included short stature, macrodontia, anteverted nares, wide nasal bridge, wide nasal base, thick eyebrows, synophrys and hypertelorism. 72% of participants had gastrointestinal complaints. Three participants were started on growth hormone and had positive results. Behavioral issues and global developmental delays were found in most participants. Neurologic abnormalities such as poor coordination and seizures (44%), indicating the importance of early screening and prophylaxis, were also very common. We have also identified minimally reported symptoms, including recurrent sinus infections (16%), markedly prevalent hearing loss (80%), as well as previously unreported migraines (20%).

PrgmNr 3426 - Loss of function variants in *PUS7* dysregulate protein synthesis and are associated with neurodevelopmental delay, microcephaly, and hyperuricemia

View session detail

Author Block: **K. Garcia**[1,2], T. Han[1,2], A. Kim[1,2], E. Macnamara[2], L. A. Schimmenti[3], E. R. Nicoli[1,2], M. Malicdan[1,2], W. A. Gahl[1,2], C. J. Tifft[1,2]; [1]Natl. Human Genome Res. Inst., NIH, Bethesda, MD, [2]Undiagnosed Diseases Program, Office of the Director, NIH, Bethesda, MD, [3]Clinical Genomics, Mayo Clinic, Rochester, MN

Disclosure Block: **K. Garcia:** None.

Protein translation is a tightly regulated process essential to maintaining homeostasis in all cells and influences gene expression. Modulation of protein translation requires coordination of various translational factors and ribosomal units, which utilize transfer RNA (tRNA) during protein synthesis. Pseudouridine synthase 7 (*PUS7*) catalyzes the conversion of uridine to pseudouridine on many RNAs. Among its targets are residues of tRNAs needed for the function of tRNA fragments (tRFs). tRFs regulate protein translation by inhibiting formation of translation initiation complex, eIF4F. Without *PUS7*, protein synthesis is dysregulated, resulting in a net increase in protein translation. Elevated protein synthesis is a hallmark of cancer and has also been implicated in autism spectrum disorder. Two patients referred to the NIH through the Undiagnosed Diseases Program presented with a neurological disease displaying a phenotype that included postnatal microcephaly, aggressive/self-injurious behavior, global developmental delay, short stature, and hyperuricemia. Although the patients exhibited clinical overlap with Lesch-Nyhan Syndrome, they did not carry pathogenic variants in *HPRT1*. Genome sequencing revealed the siblings were compound heterozygous for novel pathogenic variants in PUS7. The paternal allele carries a missense variant with no effect on protein expression, stability, or localization, but the variant is found within the catalytic domain. The maternal allele carries a splice variant that disrupts the native donor splice site causing mis-splicing of mRNA, which introduces a frameshift mutation and premature termination, resulting in degradation of the transcript via nonsense-mediated mRNA decay. We demonstrated that loss of *PUS7* activity in patient fibroblasts is sufficient to disrupt protein translation and replicated those findings in a *PUS7* knockout HeLa cell line. We then investigated the expression of genes that are potentially altered by dysregulated protein synthesis. We found that patient fibroblasts express elevated levels of MYC protein without increasing cell proliferation. Concurrently, dysregulation of protein synthesis in patient fibroblasts also decreased protein levels of HPRT1, potentially explaining the phenotypic overlap with Lesch-Nyhan Syndrome. We believe that these findings strengthen the correlation between dysregulation of protein translation and complex neurodevelopmental diseases and that the mechanism resembles those previously linked to autism spectrum disorders.

PrgmNr 3427 - Maternal mosaicism for a missense variant in the *SMS* gene that causes Snyder-Robinson syndrome

View session detail

Author Block: **M. Marhabaie**[1], S. E. Hickey[2], K. Miller[1], O. Grischow[1], K. Schieffer[3], S. Franklin[1], D. Gordon[4], S. Choi[1], P. White[1], D. C. Koboldt[2], R. K. Wilson[5]; [1]Nationwide Children's Hosp., Columbus, OH, [2]Nationwide Children's Hosp, Columbus, OH, [3]Nationwide Children's Hosp., Dublin, OH, [4]The Inst. for Genomic Med. at Nationwide Children's Hosp., Columbus, OH, [5]Nationwide Childrenâ□□s Hosp. Inst. for Genomic Med., Columbus, OH

Disclosure Block: **M. Marhabaie:** None.

There is increasing recognition for the contribution of genetic mosaicism to human disease, particularly since the advent of high-throughput sequencing which enables detection of variation at very low allele frequencies. Here, we describe an infant male who presented at nine months of age with hypotonia, dysmorphic features, congenital heart disease, hyperinsulinemic hypoglycemia, hypothyroidism, and bilateral sensorineural hearing loss. Extensive biochemical and genetic testing - including clinical whole-exome sequencing - failed to provide a diagnosis. The family was enrolled into an IRB-approved research study for trio whole-genome sequencing, which uncovered an apparent *de novo* mutation in the X-linked *SMS* gene. *SMS* encodes spermine synthase, which catalyzes the production of spermine from spermidine. Inactivation of *SMS* gene disrupts the spermidine/spermine ratio, resulting in Snyder-Robinson syndrome. The variant in our patient is absent from the gnomAD database and causes a missense change (p.Arg130Cys) predicted to be damaging by most *in silico* tools. Further, this amino acid change has been described in one patient in the literature and is believed to reduce SMS dimer stability by disrupting the structure of the adjacent spermine binding site. While Sanger sequencing confirmed the *de novo* status in our proband, PCR and deep targeted resequencing to >85,000x depth revealed that the variant is present in the maternal blood at ~3% variant allele frequency. Our findings thus provided a long-sought diagnosis for the family while highlighting the role of parental mosaicism in severe genetic disorders.

PrgmNr 3428 - Novel *PRUNE1* c.933G>A (p.Thr311=) synonymous splice variant induces exon 7 skipping and leads to an atypical presentation of NMIHBA syndrome: Case report and review of the literature

View session detail

Author Block: **C. Magyar**[1,2,3,4], D. R. Murdock[5], L. C. Burrage[6], H. Dai[5], S. R. Lalani[5], R. A. Lewis[7,8], Y-Z. Lin[9], M. Astudillo[9], J. A. Rosenfeld[5], M. Alejandro[5], J. B. Gibson[10], Undiagnosed Diseases Network, C. A. Bacino[11], B. Lee[12], H-T. Chao[3,4,5,13,14]; [1]Graduate Program in Genetics and Genomics, Baylor Coll. of Med., Houston, TX, [2]Med. Scientist Training Program, Baylor Coll. of Med., Houston, TX, [3]Jan and Dan Duncan Neurological Res. Inst., Houston, TX, [4]McNair Med. Inst., The Robert and Janice McNair Fndn., Houston, TX, [5]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [6]Dept. of Molecular and Human Genetics, Baylor Coll. Med., Houston, TX, [7]Baylor Col Med, T-803 Taub Building, Houston, TX, [8]Dept. of Ophthalmology, Baylor Coll. of Med., Houston, TX, [9]Dept. of Pediatrics, Section of Diabetes and Endocrinology, Texas Childrenâ□□s Hosp., Baylor Coll. of Med., Houston, TX, [10]Dell Children's Med. Group, Austin, TX, [11]Baylor Coll. of Med., Houston, TX, [12]Baylor Coll. Med., Houston, TX, [13]Dept. of Pediatrics, Section of Neurology and Dev.al NeuroSci., Baylor Coll. of Med., Houston, TX, [14]Dept. of NeuroSci., Baylor Coll. of Med., Houston, TX
Disclosure Block: **C. Magyar:** None.

Prune exopolyphosphatase 1 (*PRUNE1)* encodes the PRUNE protein, a phosphodiesterase member of the DHH (aspartic acid-histidine-histidine) phosphodiesterase superfamily, which plays a pivotal role in cell migration and proliferation via interactions with beta-tubulin and is highly expressed in the developing fetal brain. In 2015, biallelic *PRUNE1* loss-of-function variants were identified to cause an autosomal recessive neurodevelopmental disorder characterized by delayed development, microcephaly, central hypotonia with peripheral spasticity, and brain structural abnormalities including thinning of the corpus callosum, cerebral and cerebellar atrophy, variable white matter disease, and delayed myelination. Other prevalent co-morbid findings included seizures, visual anomalies, and gastrointestinal complications. This condition is now known as the Neurodevelopmental Disorder with Microcephaly, Hypotonia, and Variable Brain Abnormalities (NMIHBA, OMIM#617481), and to date 46 individuals have been reported in the literature. However, the phenotypic spectrum of *PRUNE1*-related disorders and their causative variants remains to be characterized fully. Here, we report a case involving a novel homozygous synonymous splice site *PRUNE1* NM_021222.1:c.933G>A (p.Thr311=) variant identified in a five-year old male from a consanguineous union with intellectual and developmental disabilities, hypotonia, and spastic diplegia, but with the absence of microcephaly, brain structural alterations, seizures, or ophthalmologic anomalies. This clinically relevant variant was identified after reanalysis of previously reported negative clinical whole exome sequencing (WES). Fragile X, *SLC16A2* sequencing, and oligonucleotide comparative genomic hybridization (CGH) array testing were unremarkable. Fibroblast RNA-sequencing revealed that the homozygous *PRUNE1* NM_021222.1:c.933G>A (p.Thr311=) variant generated an alternate transcript with an in-frame skipping of exon 7. In contrast to the previously identified *PRUNE1* c.521-2A>G splice acceptor variant, the resulting transcript in our proband is predicted to generate a mutant protein. This case represents the first synonymous splice site variant and the third pathogenic variant known to date affecting the DHH-associated domain (DHHA2 domain). These findings extend the genotypic and phenotypic spectrum in *PRUNE1*-related disorders and highlight the importance of considering synonymous and splice site variants in atypical presentations of a known disorder.

PrgmNr 3429 - Role of *VRK1* defects in spinal muscular atrophies like patients with negative *SMN1* finding

View session detail

Author Block: **C. Wu**[1,2], X. Zhao[1,2], D. Castro[3,4], K. Batley[3,4], D. Michelson[5], C. Eng[1,2], L. Meng[1,2], H. Dai[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics, Houston, TX, [3]Univ. of Texas Southwestern, Dallas, TX, [4]Childrenâ□□s Hlth.Dallas, Dallas, TX, [5]Loma Linda Univ. Med. Ctr., Loma Linda, CA
Disclosure Block: **C. Wu:** None.
Introduction: Historically, defects in *VRK1* were thought to cause autosomal recessive pontocerebellar hypoplasia type 1A (MIM# 607596), a rare form of spinal muscular atrophies (SMA) characterized by severe neurodegenerative disorders affecting growth and function of the brainstem and cerebellum, resulting in little or no development. To date, fewer than 15 cases of *VRK1*-related conditions have been reported in the literature. The spectrum of symptom severity and age of onset are still largely unknown. **Methods:** Experimental procedures and bioinformatics analysis of exome sequencing (ES) were performed following previously described methods. Variant interpretation was performed according to the joint American College of Medical Genetics and Genomics and Association for Molecular Pathology guidelines. **Results**: Patient 1 was a boy who presented in early childhood with progressive limb weakness and loss of motor function, microcephaly, and scoliosis and he progressed in adolescence to severe dysphagia, dysarthria, and respiratory insufficiency. *SMN1* deletion and *IGHMBP2* deletion/duplication tests were negative. ES identified compound heterozygous likely pathogenic variant c.266G>A (p.R89Q) and pathogenic variant c.706G>A (V236M) in *VRK1*. Patient 2 was a 14 yo female presented with peripheral neuropathy, pes cavus, joint hypermobility, and scoliosis. She had three similarly affected sisters (patients 3, 4, and 5, respectively). All sisters had normal cognition, with normal imaging of the brain confirmed in patients 2 and 3. They were all homozygous for c.706G>A (p.V236M) in *VRK1* inherited from the heterozygous parents. Patient 6 was a previously reported 32 yo male with progressive extremity weakness, breathing and swallowing difficulty. He was clinically diagnosed with early-onset amyotrophic lateral sclerosis (ALS). Testing for *SMN1* deletion and Kennedy disease were negative. ES identified compound heterozygous likely pathogenic variants c.356A>G (p.H119R) and c.961C>T (R321C) in *VRK1*. This case represents the mildest presentation among all reported patients with *VRK1*-related conditions. Pontocerebellar hypoplasia was not a consistent finding in these six patients. **Conclusion:** We have presented six patients from three unrelated families caused by biallelic variants in *VRK1*, which demonstrates a wide phenotypic spectrum ranging from muscle weakness to adult-onset ALS. Progressive muscle weakness can begin in different ages, with variable severity, resembling the spectrum of *SMN1*-associated SMA. These results highlighted the importance of *VRK1* evaluation, even when neuroimaging does not suggest pontocerebellar hypoplasia.

PrgmNr 3430 - Role of exome testing in autism spectrum disorder

View session detail

Author Block: **J. Castorena Ibarra**[1], R. Villafuerte[1], D. Aguilar[2]; [1]Escuela de Mediciana y Ciencias de la Salud del Tecnologico de Monterrey, Monterrey, Mexico, [2]Hosp. Zambrano Hellion Tec Salud, San Pedro Garza GarcÃa, Mexico
Disclosure Block: **J. Castorena Ibarra:** None.
Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder. Recently numerous exome sequencing studies had revealed new genes associated with ASD including de novo and hereditary mutations. TAF1 Transcription initiation factor TFIID subunit 1, also known as TATA-box binding protein associated factor 1 gene is critical for the normal function of neurons in the brain. TAF1 provides instructions for making part of a protein called transcription factor IID (TFIID). Recently a new X-linked syndrome (MIM# 300966) characterized by intellectual disability, autistic behaviors and distinctive facial features have been identified in association with pathogenic variants (PV) in TAF1. To date less than 40 cases have been described in 27 families, without including hispanic population.The objective of this abstract is to describe a mexican male infant with ASD and dysmorphias in which exome testing reveal a missense variant in TAF1.Case description: A 2 years 11 months male patient from healthy non-consanguineous parents and without relevant prenatal history. Geneticist evaluation reveals delayed speech and language development as well as craniofacial dysmorphic features. Under clinical suspicion of syndromic intellectual disability, exome testing was required in which a missense variant in TAF1 gene c.2189G>A p.Gly730Glu located in exon 14 was identified. Results: Patient's dysmorphic and behavioral features were compared with those previously described in Chang et al. and Oâ□□Rawe et al in TAF1 PV carriers, showing the most prevalent including global developmental and speech and language delay (95%), motor delay (50%) and hypotonia (70%). Autistic behavior (16%), hypertelorism, retrognathia, bulbous nose (20%) and wide nasal bridge (12.5%). To date 34 PVÂ´s have been described located throughout the gene. Missense variant identified in case described (c.2189G>A p.Gly730Glu) located in exon 14 encodes for an evolutionarily conserved central domain (DUF3591). Previously OÂ´Rawe and Cheng reported 10 PVÂ´s in TAF1 falling within domain DUF3591 which encompasses the TAF1 HAT domain and numerous points of contact with TAF7.The TAF1 missense variant identified by exome sequencing is not present in population databases nor has been reported in the literature. According to MutationTaster and at least five prediction models more, it is predicted to be a damaging or disease-causing variant but according to SIFT and other prediction models is classified as well tolerated or benign variant. However, conflictive interpretation, clinical data strongly suggest X-linked syndrome associated with TAF1 diagnosis.

PrgmNr 3431 - The phenotypic trajectory of 16p12.1 deletion is determined by family history of disease

View session detail

Author Block: **L. Pizzo**[1], C. Smolen[1], M. C. Jensen[1], A. Tyryshkina[1], L. Rohan[1], E. Huber[1], C. Taylor[2], 16p12.1 deletion Consortium, S. Girirajan[1]; [1]Pennsylvania State Univ., University Park, PA, [2]Geisinger Hlth.Systems, Lewisburg, PA
Disclosure Block: **L. Pizzo:** None.
The extreme phenotypic variability associated with copy-number variants such as the 16p12.1 deletion impacts genetic diagnosis and leads to challenges in predicting prognosis and management of affected families. Here, we recruited 146 children and 81 adults with 16p12.1 deletion from 113 families, and performed deep phenotyping through analysis of medical records, detailed clinical questionnaires, phone interviews, and quantitative assessment of cognitive and behavioral domains. We observed that probands manifested a range of phenotypes, including speech delay (73%), developmental delay/intellectual disability (63%), ASD (44%), ADHD (39%), craniofacial features (30%), hypotonia (30%) and multiple other congenital features. Further, probands with nervous system defects were more likely (OR=5.5) to present comorbid psychiatric and behavioral phenotypes (FDR=0.01). In contrast, adults with the deletion showed depression and anxiety (52%), learning difficulties in school (41%), schizophrenic features or bipolar disorder (14%), intellectual disability (11%), and alcohol and/or drug addiction (9%). Moreover, probands presented more severe distributions of behavioral functioning (4.89 SD), communication skill (2.44 SD), and non-verbal IQ (2.25 SD) scores compared to their carrier parents, comparable to those observed among individuals with autism from the Simons Simplex Collection. Next, we studied whether parental phenotypes could predict the phenotypic manifestation in children. K-means clustering of parental phenotypes showed that probands (n=38) with parents who presented cognitive (80%), schizophrenic (64%), depression and anxiety (64%), or addiction features (32%) were more likely to present defects in behavioral (p=0.04) and psychiatric (p=0.03) domains, compared to probands (n=25) whose parents had milder psychiatric features (8-15%). To understand the contribution of family history and rare variants towards phenotypic manifestation, we performed whole-genome sequencing and SNP array analyses of affected families. We observed that probands with a strong family history of psychiatric phenotypes presented a higher burden of loss-of-function mutations in genes intolerant to variation (p=0.01), while no differences were observed for likely deleterious missense mutations, deleted, or duplicated genes. Our work highlights the relevance of family history towards predicting behavioral and psychiatric outcomes in children with 16p12.1 deletion, and suggests that a subset of rare variants contribute towards more severe clinical manifestation of the deletion in the presence of strong family history of disease.

PrgmNr 3432 - The Use of Acetazolamide for Apneic Episodes in a Patient with EEF1A2 Mutation: A Case Report

View session detail

Author Block: **C. Joyce**[1], J. Sheah[2], j. n. peeden[3]; [1]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [2]Dept. of Pediatric Neurology, East Tennessee Children's Hosp., Knoxville, TN, [3]East Tennessee Children's Hosp./Univ. of Tennessee Dept Med., Knoxville, TN
Disclosure Block: **C. Joyce:** None.
OBJECTIVE: Recurrent and significant daytime apnea events resulting in cyanosis and oxygen desaturations were quite problematic in a 6-year-old female with global neurodevelopmental delay due to an EEF1A2 mutation. After many trials the patient was found to be responsive to Acetazolamide in treatment of the apneic episodes.
METHODS: A 6-year-old female presented with recurrent apenic episodes occurring only while awake. Medical evaluation included continuous electroencephalogram (EEG) with events captured, pH probe also with events captured, pulmonary evaluation, magnetic resonance imaging (MRI) of the brain and echocardiogram. These diagnostic tests failed to identify any specific cause of the cyanotic events. Multiple medication regimens were attempted throughout the evaluation period including levetiracetam, baclofen, clonidine, oxcarbazepine, glycopyrrolate, and scopolamine all of which resulted in no improvement of cyanotic events. After secondary causes of apnea had been excluded, acetazolamide was prescribed given a favorable safety profile and case reports of success in Pitt-Hopkins syndrome for similar daytime apneic events. This is the first report of this therapy improving the intensity and frequency of apenic episodes in patients with the EEF1A2 mutation.
RESULTS: Acetazolamide was used in an attempt to ameliorate cyanotic events in a patient with EEF1A2 mutation with success. There is currently no literature in regards to patients with this mutation and the use of acetazolamide.
CONCLUSIONS: This is the first report of a patient with an EEF1A2 mutation responding to Acetazolamide for treatment of day-time apneic episodes.

PrgmNr 3433 - Triple knockout of*Fmr1,Fxr1*and*Fxr2*in Purkinje neurons results in mild behavioral phenotypes and a dramatic reduction of testis weight

View session detail

Author Block: **Y. Gu**, D. L. Nelson; Baylor Coll. of Med., Houston, TX
Disclosure Block: **Y. Gu:** None.

Fragile X syndrome (FXS) is a common cause of intellectual disability and autism due to mutation of *FMR1*. *Fmr1*knockout mice recapitulate major phenotypes of FXS patients and offer a very useful model for studying the function of the FMR1 product, FMRP. Although tremendous advances have been made in understanding FXS in the past several decades, effective treatment of FXS patients remains elusive.*FMR1*has two related paralogs, *FXR1*and*FXR2* which share significant homology and common functional domains. We hypothesize that phenotypes in FXS patients and *Fmr1* knockout mice are altered through compensation by the paralogs, masking the effects of loss of FMRP. Double knockout mice for *Fmr1* and *Fxr2* showed exacerbated behavioral abnormalities and disturbed lipid and glucose metabolism. These phenotypes are absent in either single knockout model, supporting our hypothesis. We recently created a second *Fxr2*knockout model (*Fxr2*KO2) that is a true null along with a conditional version of this allele (*Fxr2*cKO). Global double knockout with *Fmr1*leads to a more severe phenotype: embryonic lethality. Global knockout of *Fxr1*is lethal in the early postnatal period. To study complementation of function in neurons, we developed tissue specific knockout of all three genes directed to Purkinje neurons using L7-Cre and conditional alleles for all three genes. One breeding scheme produced male mice that lacked *Fmr1*and were heterozygous for Fxr1 and Fxr2 in Purkinje neurons and controls lacking the Cre (*Fmr1*cKO/y; *Fxr1*cKO/+; *Fxr2*cKO/+). Several behavioral tests (rotarod, parallel rod foot slip, digital gait, metal mesh flip and mini motor) demonstrated only minor differences. Using a different mating strategy, we produced and analyzed male mice with three genotypes: *Fmr1*cKO/y; *Fxr1*cKO/cKO; *Fxr2*cKO/cKO with or without L7-Cre and *Fmr1*cKO/y; *Fxr1*cKO/+; *Fxr2*cKO/cKO with L7-Cre. Surprisingly, mice lacking all three genes in Purkinje neurons were only mildly affected, they covered a shorter distance at a slower speed than the other two groups in the parallel rod footslip test. Significant differences in several other parameters were present, but Purkinje neuron triple KO mice were not ataxic. Unexpectedly, testis weight in the Purkinje neuron triple KO mice was reduced by ~ 35% compared with other two groups. Previous studies, along with macroorchidism in FXS have suggested a role for these proteins in testis development. These results demonstrate that L7-Cre expression is not confined to Purkinje neurons. It is noteworthy that neurons lacking all three proteins are viable and functional, suggesting alternative models for compensation of function within this gene family.

PrgmNr 3434 - Variation of Serum Metabolome and Exosomal Transcriptome in Acute Ischemic Stroke: Results of the Metabolome in Ischemic Stroke Study (MISS)

View session detail

Author Block: **E. Sidorov**[1], C. Xu[2], S. Goyal[3], J. Garcia[4], C. Seraphin[4], A. Blair[4], D. Gordon[4], J. Chainakul[4], D. K. Sanghera[3]; [1]Oklahoma Univ. of Hlth.Sci. Ctr., Oklahoma, OK, [2]Dept. of Biostatistics and Epidemiology, Oklahoma Univ. of Hlth.Sci. Ctr., Oklahoma, OK, [3]Dept. of Pediatrics, Coll. of Med., Oklahoma Univ. of Hlth.Sci. Ctr., Oklahoma, OK, [4]Dept. of Neurology, Coll. of Med., Oklahoma Univ. of Hlth.Sci. Ctr., Oklahoma, OK
Disclosure Block: **E. Sidorov:** None.
Stroke is the second-leading cause of death worldwide and a major cause of serious disability in the US. Oklahoma has the 9th highest rate of stroke in the nation among all states. Previous studies found multiple associations of serum metabolites with acute ischemic stroke (AIS) against controls, but very few studies have evaluated serum metabolites during the acute and chronic phases of AIS. The exosomal RNAs/miRNAs have emerged as biomarkers and therapeutic targets for various diseases, however, their function in stroke remains largely unknown. In this study, we compared the metabolomic patterns of acute and chronic stages of ischemic stroke. In addition, in a subset of individuals, we have also evaluated changes in the serum exosomal transcriptome during the acute and chronic stage of AIS by performing RNA sequencing. We evaluated 1295 serum metabolites on 60 stroke patients at acute and chronic stages of stroke by performing global metabolomics using ultra-high-performance liquid chromatography/mass spectrometry (LC-MS) and gas chromatography/mass spectrometry (GC-MS) at Metabolon Inc. Orthogonal Partial Least Square-Discrimination Analysis (OPLS-DA) and Principal Component Analysis (PCA) were used to inspect group disparity using multivariate data. The OPLS-DA revealed significant separation of acute and chronic stage metabolites, PCA showed no hidden structures. Exosomal miRNA hsa-miR-9-3p and hsa-miR-129-5p were observed to be differentially expressed showing 14.7- and 7.6-fold increase with FDR p-value of $2.83x10^{-6}$ and $2.87x10^{-3}$, respectively, at the acute phase compared to the chronic phase of stroke. Correlation analysis of miRNAs with metabolites revealed ceramides (d18:1/17:0, d17:1/18:0)* and (d18:2/24:1, d18:1/24:2)* and sphingomyelin (d18:2/18:1)* to be strongly correlated with miR-129-5p (r= 0.78-0.84; p

PrgmNr 3435 - A heterozygous gain-of-function variant in *KIF5B* causes osteogenesis imperfecta by disrupting the Golgi-primary cilia axis

View session detail

Author Block: **M. Washington**[1], B. Zhang[2], I. Song[1], L. C. Burrage[3], V. Rossi[1], M. Nonet[2], J. Chen[2], D. Baldridge[2], G. A. Silverman[4], V. R. Sutton[3], J. A. Rosenfeld[1], M. J. Hicks[1], D. R. Murdock[1], H. Dai[1], M. Weis[5], Undiagnosed Diseases Network, D. R. Eyre[5], S. C. Pak[2], T. Schedl[6], B. Lee[3], R. Marom[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Washington Univ. Sch. of Med., St. Louis, MO, [3]Baylor Coll. Med., Houston, TX, [4]Washington Univ. Sch. of Med., St. Louis, MO, [5]Univ. of Washington, Seattle, WA, [6]Washington Univ. Sch. of Med., St Lous, MO

Disclosure Block: **M. Washington:** None.

Kinesin superfamily motor proteins (KIFs) play an essential role in fundamental cellular processes, including transport of secretory vesicles, endocytosis, organelle trafficking, cell cycle and ciliogenesis. Pathogenic variants in KIF-related genes have been implicated in microcephaly, neurodevelopmental disorders, and skeletal dysplasia. Exome sequencing identified a *de novo*, heterozygous variant, c.260C>T (p.Thr87Ile) in *KIF5B*, encoding the kinesin-1 heavy chain, in an individual with osteogenesis imperfecta (OI) of moderate severity. The patient presented with multiple fractures, low bone mineral density, short stature, scoliosis, dentinogenesis imperfecta, and mild hearing loss. Genetic testing for known OI genes was negative. The *KIF5B* variant substitutes a highly conserved amino acid residue within the P-loop of the kinesin motor domain. Other missense mutations in Thr87 were shown to destabilize the P-loop structure and interfere with ADP binding *in vitro*. To better understand the *in vivo* genetic mechanism of *KIF5B* p.Thr87Ile we modeled the variant in the C. elegans ortholog, unc-116, at the corresponding position (Thr90Ile) by CRISPR/Cas9 editing. The *unc*-116 Thr90Ile homozygotes were L1 larval lethal. Furthermore, variant heterozygotes displayed dominant motility and body morphology phenotypes that were not observed in *unc-116* deletion heterozygotes. Together, these data suggest that the *KIF5B* variant is damaging and displaying dominant behavior that is distinct from loss of function. To further assess the cellular consequences of the *KIF5B* p.Thr87Ile variant we performed a microscopy study in primary patientâ□□s fibroblasts and NIH3T3 cells overexpressing the human *KIF5B* mutation. Electron microscopy in patient fibroblasts showed dilated endoplasmic reticulum with granular material, and multiple vacuoles containing cell breakdown products. Immunofluorescence labeling of the Golgi matrix protein, GM-130 revealed abnormal distribution of the Golgi complex. Overexpression of mutated *KIF5B* in NIH3T3 cells was associated with elongation of the primary cilia, as assessed by immunofluorescent staining for acetylated tubulin. Notably, overexpression of the wild type *KIF5B* showed a similar phenotype, while KIF5B depletion using siRNA did not significantly alter the cilia length. Our data support that the *KIF5B* p.Thr87Ile variant is pathogenic and acts via a gain-of-function mechanism to disrupt the secretory pathway and Golgi-primary cilia axis. This study expands the spectrum of KIF-related disorders and emphasizes the critical role of *KIF5B* in skeletal development.

PrgmNr 3436 - Biallelic variants in *LDB3*: A new inheritance pattern for an old cardiomyopathy gene

View session detail

Author Block: **G. Gotway**[1], N. Baez Hernandez[1], C. Uhles[2]; [1]Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [2]Children's Med. Ctr. Dallas, Dallas, TX
Disclosure Block: **G. Gotway:** None.
The *LDB3* gene codes for a PDZ-domain protein with multiple splice forms that localizes to Z bands in striated muscle. Heterozygous variants in *LDB3* were reported in patients with autosomal dominant cardiomyopathy, especially left ventricular noncompaction; however subsequent studies cast doubt on the relation of heterozygous variants to the disease. Knockout models in mice demonstrated that biallelic complete loss of function of Ldb3 causes neonatal death due to cardiomyopathy. However, specific knockout of the long forms of Ldb3 leads to a milder phenotype, with early cardiomyopathy and partial lethality, then improvement in survivors with recurrence in adulthood. We identified a pair of brothers who both developed left ventricular noncompaction cardiomyopathy as newborns, which subsequently improved in both cases. Molecular sequencing of genes associated with cardiomyopathy demonstrated that the only variants they had in common were two variants in trans in the *LDB3* gene. The first variant, c.655C>T (p.R219X), is expected to cause an early truncation that would lead to loss of all isoforms of LDB3. The second variant, c.1798C>T (p.R600X), is predicted to lead to loss of only the long isoforms of the protein. Thus, the variants in these brothers are expected to have the same effect on protein expression of LDB3 as that seen in the mouse model deficient for the long forms of Ldb3. According to gnomAD, variants leading to specific loss of the long forms of LDB3 have a carrier frequency of about 1 in 3,000; variants leading to loss of all isoforms are about half as common. Therefore while variants causing loss of function are too common to cause highly penetrant autosomal dominant disease, the frequency is consistent with these variants being a rare cause of autosomal recessive cardiomyopathy. Given the similarity between our patients and the mouse models of loss of Ldb3, we propose that biallelic variants in *LDB3* cause autosomal recessive cardiomyopathy. Specific loss of the long forms of LDB3 leads to a milder phenotype with neonatal cardiomyopathy that resolves, but could potentially recur at a later time.

PrgmNr 3437 - Combined PMM2-CDG and hereditary fructose intolerance in a child with mild clinical presentation: a case report

View session detail

Author Block: **X. Hong**[1], M. He[2], A. C. Edmondson[1]; [1]Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Childrenâ□□s Hosp. of philadelphia, Philadelphia, PA
Disclosure Block: **X. Hong:** None.
The patient, who was born to consanguineous parents, first came to medical attention at 6 months of age for failure to thrive in the setting of hepatomegaly with transaminitis, hypotonia, and esotropia. A comprehensive workup was most notable for abnormal carbohydrate deficient transferrin (CDT) study, suggesting a Type I Congenital Disorder of Glycosylation (CDG). Whole exome sequencing revealed homozygous pathogenic variants, c.448G>C (p. A150P), in *ALDOB*, establishing a diagnosis of Hereditary Fructose Intolerance (HFI); and homozygous likely pathogenic variants, c.44G>C (p. G15A), in *PMM2*, concerning for PMM2-CDG. Phosphomannomutase 2 (PMM2) activity was reduced in leukocyte and plasma N-glycan analysis revealed an under-mannosylation profile, indicating functional GDP-mannose deficiency. The diagnosis for PMM2-CDG was thus confirmed.

HFI is caused by deficient aldolase B, which reversibly catabolizes fructose-1-phosphate (Fru-1-P) and fructose-6-phosphate (Fru-6-P). Uncontrolled HFI patients are at risk for hypoglycemia, liver dysfunction and metabolic decompensation and can display Type I CDT patterns, as the accumulated Fru-1-P inhibits GDP-mannose biosynthesis. HFI can be treated effectively with dietary restriction of fructose.

PMM2-CDG is the most common CDG subtype. PMM2 converts mannose-6-phosphate (Man-6-P) to mannose-1-phosphate (Man-1-P), precursor of GDP-mannose. Its deficiency leads to hypoglycosylation of numerous glycoproteins, transferrin included. The clinical presentation of PMM2-CDG varies greatly from severe antenatal presentation with multisystem involvement to mild adulthood presentation. Despite extensive ongoing research, management of PMM2-CDG is still supportive.

The patient had minimal fructose exposure prior to diagnosis and fructose-restricted diet was initiated afterward. Persistent but slowly improving, Type I CDT patterns were observed in the following 4 years. Spontaneous normalization of hypoglycosylated transferrin regardless of clinical presentation has been described for PMM2-CDG. Persistent under-mannosylation N-glycan profiles were also observed, typical for PMM2-CDG and MPI-CDG patients but may also for HFI patients in early infancy. Interestingly, this profile was not detected in two older fructose-restricted HFI patients.

Overall, our patient with combined PMM2-CDG and HFI has a low severity score based on the Nijmegen Pediatric CDG Rating Scale. It may be due to the high residual PMM2 activity from the private *PMM2* variant, p. A150P. Alternatively, the deficiency in aldolase B may compensate PMM2 deficiency through a mechanism yet to be delineated.

PrgmNr 3438 - Implementing genome and transcriptome sequencing methods to improve the diagnosis of Mendelian myopathies

View session detail

Author Block: **V. Ganesh**[1,2], B. Weisburd[1], F. Aguet[1], G. Tiao[1], H. L. Rehm[3,1], A. O'Donnell-Luria[1,4]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Brigham and Women's Hosp., Boston, MA, [3]Massachusetts Gen. Hosp., Boston, MA, [4]Boston Children's Hosp., Boston, MA
Disclosure Block: **V. Ganesh:** None.
Approximately two-thirds of patients with Mendelian myopathies lack a specific molecular diagnosis. Although the adoption of gene panel and exome sequencing in clinical practice improved the diagnostic sensitivity in these diseases, the ability to detect rare variants outpaces our ability to interpret their functional consequences. Transcriptome sequencing (RNA-seq) emerged as a useful technology to improve variant resolution and diagnostic yield in rare diseases. Prior work in the Broad Center for Mendelian Genomics was the first to demonstrate the utility of RNA-seq for genetic diagnosis in Mendelian myopathies. We expand upon this tool by incorporating autoencoder denoising methods (FRASER, OUTRIDER) paired with a web browser visualization tool to identify outlier gene expression and splice variants in a cohort of muscle biopsies from 113 individuals with suspected Mendelian myopathies. Implementation of these methods 1) improved the diagnostic yield in our cohort by 13%, 2) validated the transcriptional effect of variants predicted to affect splicing, and 3) facilitated the identification of new gene-disease associations. This demonstrates the power of RNA-seq in the diagnosis of Mendelian myopathies and provides a method to validate suspected transcriptional effects of noncoding variants. It also demonstrates an improvement in RNA-seq analysis specificity by controlling for latent confounders using autoencoder-based methods.

PrgmNr 3439 - Intravenous fish oil emulsion and fatty acid profiles

View session detail

Author Block: **P. Tanpaiboon**[1], D. Salazar[2], C. Kucera[1], B. Epstein[1], J. Taylor[1], R. Sharma[1], F. L. Lacbawan[2]; [1]Quest Diagnostics, San Juan Capistrano, CA, [2]Quest Diagnostics Nichols Inst, San Juan Capistrano, CA
Disclosure Block: **P. Tanpaiboon:** Salary/Employment; Quest Diagnostics.
Fish oil triglyceride injectable emulsion (OmegavenÂ®) is used as a source of calories and fatty acids in pediatric patients with parenteral nutrition-associated cholestasis (PNAC). This fish oil-based intravenous fat emulsion (FO-IVFE) contains high omega-3 derivatives, eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA), with high total omega-3 fatty acid (30%-60% of total fatty acids). The percent of composition of EPA and DHA by weight (13%-27%) is higher than soybean-based IVFE (0%). While managing patients with FO-IVFE, fatty acids (FA) status should be monitored regularly to determine the efficacy of treatment and identify essential fatty acid (EFA) deficiency. EFA profiles of patients receiving this product have been reported in several studies; as expected, levels of EPA and, to a lesser degree DHA are markedly elevated. However, the effect of this product on serum levels of branched-chain saturated FA such as phytanic and pristanic acids is less well characterized. At least one study demonstrates the effect of FO-IVFE on levels of branched-chain FA and EFA in preterm infants. We observed characteristic FA profiles including elevated omega-3 derivatives, elevated pristanic and phytanic acids, and decreased linoleic acid in several pediatric patients. Clinical information was available in 12 patients, ages 2 months to 10 years old. All patients received FO-IVFE when the samples were collected. EPA, DHA, total omega-3, and pristanic acids were elevated in all patients. EPA and DHA were the most prominent (10-26 times and 1-4 times the upper limit of normal (UNL), respectively). Pristanic acid levels were (2-20 times UNL) correlated with EPA levels. Phytanic acid levels were elevated to a lesser degree (0-4 times UNL) compared to pristanic acid. Linoleic acid levels were low in all 12 patients. While mild elevations of phytanic and pristanic acids can be seen secondary to diets rich in animal fats or dairy products, moderate to marked elevations are more often associated with peroxisomal disorders. Consequently, FA analysis in patients treated with FO-IVFE could lead to diagnostic pitfalls in some cases, particularly when only very long chain fatty acid (VLCFA) analysis is performed following a positive newborn screen, where the accompanying increases of omega-3 fatty acids would not be reported. In the absence of clinical information, elevated levels of pristanic and phytanic acids in a limited VLCFA panel could be falsely suggestive of an underlying peroxisomal defects. It would be prudent to mention FO-IVFE treatment in the clinical interpretation for any VLCFA sample in which phytanic and pristanic acids are significantly elevated.

PrgmNr 3440 - Lessons learned from a severe Isovaleric acidemia (IVA) patient from hyperammonemia

View session detail

Author Block: **C-H. A. Tsai**, B. Jessica, m. Chou, c. Zimmerman, D. DeMarzo, Y. Enchautegui-Colon; Children's Hosp. Oklahoma, Oklahoma City, OK
Disclosure Block: **C.A. Tsai:** None.
With the execution of expanded newborn screen program nationwide, it is uncommon to see severe hyperammonemia associated with IVA. We present a seven-day-old boy with severe IVA complicated with hyperammonemia. This child was flagged by NBS at day 4, but due to COVID19 and parental skepticism, confirmatory testing was delayed and a leu restricted diet was not followed as instructed. The patient initially presented to the ED with Ammonia of 588 ug/dL which then increased to 1000 ug/dL. This child received carnitine, arginine, carglumic acid (Carbaglu), CRRT and Ammonul. **The Ammonul was not helpful.** Ammonul has two major components sodium benzoate and sodium phenylacetate. Sodium benzoate functions by removing glycine with ammonia which could be harmful for this condition as IVA needs to bind to glycine as a rescue. Sodium phenylacetate binds to glutamine to remove ammonia. However, this patientâ□□s plasma amino acid assay revealed a low glutamine level of 256umol/L. The hyperammonemia was corrected in 15 hours and with the use of Carbaglu there was no rebound of hyperammonemia. Our patient suffered from bone marrow suppression associated to the organic acidemia and required frequent multiple platelet transfusions and GCSF for neutropenia. The management of this patient provides supporting evidence of the many theoretic metabolic â□□factsâ□□ including why Ammonul is not helpful in organic acidemias.

PrgmNr 3441 - Mutational spectrum in a sample of mexican patients with Hunter syndrome

View session detail

Author Block: **M. A. Ramirez**[1,2], L. E. Figuera[2], L. d. Rizo[2], J. E. Garcia[2], S. d. Mendoza[2], M. T. MagaÃ±a[2], L. Arnaud[3], M. P. Gallegos[2]; [1]Univ. de Guadalajara, Guadalajara, Mexico, [2]Centro de InvestigaciÃ³n BiomÃ©dica de Occidente, IMSS, Guadalajara, Mexico, [3]Hosp. Civil Dr. Juan I. Menchaca, Guadalajara, Mexico

Disclosure Block: **M.A. Ramirez:** None.

INTRODUCTION: Hunter syndrome or MPS II (OMIM #309900) is an X-linked recessive disease caused by deficiency of iduronate-2-sulfatase lysosomal enzyme, which leads to progressive accumulation of heparan and dermatan sulfate in different organs and tissues. It has a global incidence of 0.69 to 1.19/100,000 live births. *IDS* is the gene involved, it has been missense mutations in 55%, nonsense in 21% and deletions in 20% of the patients. Some patients with more severe phenotypes have total deletions or rearrangements between *IDS-IDS2*. There is only one study in Mexican population in which 36% presents a complex rearrangement.OBJECTIVE: To determine the mutations and deletions frequency in *IDS* gene and IDS2 pseudogene in patients with HS.MATERIAL AND METHODS: 7 DNA samples from patients with clinical diagnosis of HS between 4 and 12 years old were analyzed. The samples were processed by multiplex ligation-dependent probe amplification with SALSA probemix P164-B2 specific for HS and analyzed by Coffalyser software. Patients with no deletion were included for Sanger sequencing for exons 2,3,4,6,7,8 and 9, the chromatograms obtained were analyzed with Chromas software.RESULTS: 28.5% (2/7) of the samples from patients with HS present a complete deletion of *IDS* gene and pseudogene *IDS2*. In the rest of the patients, mutations in intron and exon 3 and in exon 9 were found. For intron 3 the variant found was CS066287 (1/7), located 1 base prior to the exon generating an alteration of the acceptor site for the splicing zone. In exon 3 we found 2 variants NM_000202.5:c.275T>G p.[Leu92Arg] (2/7) and NM_000202.5:c.284_287del p.[Arg95Asnfs*34] (1/7), the first corresponds to and missense variant and the second corresponds to a small deletion of 4 pb with a change of the reading frame. For exon 9, NM_000202.5:c.1229_1229delT p.[Leu410Arg fs*30] was found, it corresponds to a punctual deletion of a T nucleotide.CONCLUSIONS: The frequency of deletions in *IDS* gene and *IDS2* pseudogene of 28.5% observed in these patients is higher than the previously described in the world literature but fewer than the previous report in our population. These deletions have been associated with severe phenotypes at the neurological level; in our sample, these patients have severe and moderate psychomotor delay. In patients with the mutations, we observed a severe phenotype no matter the kind of mutation found. For the other systems, none of them present alterations in ocular level, all except one present alterations in oral cavity, only 2 patients show hearing loss, but everyone presents affectation in the skeletal system, principally in the loss of functionality in hands, kyphosis/scoliosis.

PrgmNr 3442 - Mutations in COL5A1 lead to heritable thoracic aortic disease

View session detail

Author Block: **D. Guo**[1], P. Guan[1], I. C. Marin[1], P. Arnaud[2], M. J. Bamshad[3], D. A. Nickerson[4], J. Backer[5], G. Jondeau[6], C. Boileau[7], D. M. Milewicz[8]; [1]Univ. of Texas McGovern Med. Sch. at Houston, Houston, TX, [2]AP-HP, HÃ´pital Bichat, France, [3]Univ. of Washington, Seattle, WA, [4]Univ Washington Sch Med, Seattle, WA, [5]Ghent Univ. Hosp., Ghent, Belgium, [6]HÃ´pital Bichat, Paris, France, [7]Hosp Bichat-Claude Bernard- Batiment Inserm, Paris, France, [8]Univ Texas McGovern Med Sch, Houston, TX
Disclosure Block: **D. Guo:** None.
To identify novel genes for heritable thoracic aortic disease (HTAD), we developed a new machine learning method to prioritize gene identification based on signaling pathways for known HTAD genes and the topological characters in the protein-protein interaction network. The method also integrates data on both gene mutational constraints from gnomAD exome v2.1.1 and tissue special gene expression from the GTEx v8.1. Using this method, all [insert #) validated HTAD genes were identified, along with 307 additional genes, and the genes were ranked in terms of potential for causing HTAD; *FBN1* (11th) and *COL3A1* (15*th*) ranked in the top 20 and *MYLK* ranked lowest at 280th. Rare (MAF -4) and damaging variants (CADD > 20) in the top 10 genes were assessed in exome data from 445 unrelated and unsolved HTAD families, 23 trios, and 355 patients with early sporadic thoracic aortic dissection (ESTAD). *COL5A1*, the 5th ranked gene, is a known cause of classic Ehlers-Danlos syndrome (EDS; haploinsufficiency and glycine substitions in triple helical region are disease-causing). The following rare variants in *COL5A1* were identified in HTAD and ESTAD cases: (1) A *de novo* coding deletion, p.Thr104del, absent in gnomAD exomes and predicted to be disease-causing by MutationTaster; (2) three rare variants, p.Gly607Arg, p.Gly1270Ser, and Gly1315Val, that alter critical glycines at the COL5A1 triple-helical region. Two of these variants segregate with HTAD in families and one is an ESTAD case; (3) two rare variants, p.Ser1376Pro and p.Pro1566Leu, that segregate with HTAD in two families. Additionally, a recent publication identified two individuals with *COL5A1*, p.Gly1414Ala and p.Gly1144Ala, disrupting triple helical glycines in exomes from 702 patients with sporadic TAD. Previous publications of 34 patients with classic EDS and vascular diseases (8 had TAD) failed to identify any triple-helical glycine substitutions. Thus, applying machine learning to identify putative HTAD genes and assessing exome data from HTAD and ESTAD cases for rare variants in these genes has identified *COL5A1* as a possible HTAD gene, and has the potential to identify further genes for HTAD.

PrgmNr 3443 - Oculocutaneous Albinism: Compound heterozygote genotype-phenotype correlation

View session detail

Author Block: **A. Alejandro Soto**[1], J. Pascual[2], O. Quincoces[2], C. I. Hernandez[3], S. Carlo[4,5], A. S. Cornier[6]; [1]Dept. of Biology, Univ. of Puerto Rico Mayaguez Campus, Mayaguez, PR, [2]Ponce Hlth.Sci. Univ. Sch. of Med., Ponce, PR, [3]Dept. of Pediatrics, Natl. Children's Hosp./ NIH, Washington, DC, [4]Ponce Hlth.Sci. Univ., Ponce, PR, [5]Genetic Section, San Jorge Children's & Womenâ□□s Hosp., San Juan, PR, [6]San Jorge Children's Hosp. / Ponce Hlth.Sci. Univ., San Juan, PR

Disclosure Block: **A. Alejandro Soto:** None.

Oculocutaneous albinism (OCA) is a group of human autosomal-recessive hypopigmentation disorders characterized by hypopigmentation in the skin, hair, and eyes [1]. OCA1 and OCA2, two of the seven types of OCA, are caused by mutations of the *TYR* and *OCA2* genes, respectively, which are responsible for most oculocutaneous albinism. We are presenting the case of an 11-year-old male Puerto Rican patient who was first evaluated by a pediatric neurologist at age 2 *12* with a history of preauricular tag, strabismus and then diagnosed with language developmental disorder associated to motor apraxia. At age 10 years the patient was referred to Genetics clinic by ophthalmologist due to accommodative esotropia, bilateral hypermetropia, infantile nystagmus, and foveal hypoplasia. The patient was diagnosed with oculocutaneous albinism (OCA) due to *TYR* gene mutations. The diagnosis was made through an albinism gene panel. The patient is specifically diagnosed with OCA 1B, also known as â□□yellow albinismâ□□, though he still presents with hypopigmentation. OCA gene panel showed two variants which are associated with an autosomal recessive *TYR*-related condition. The first is a heterozygous likely pathogenic variant: p.Gly47Asp. The variant is predicted to result in a single amino acid substitution (missense) of glycine to aspartic acid at codon 47 in exon 1 of the *TYR* gene. A known variant at this position, p.Gly47Val, has already been associated with OCA1. A second heterozygous variant of unknown significance: p.Arg402Gln, was documented as well. This variant is predicted to result in a single amino acid substitution (missense) of arginine to glutamine, at codon 402 in exon 4 of the *TYR* gene. This variant has been reported as a common polymorphism with other mutations *in trans* in multiple individuals with autosomal recessive OCA and has also been described as a hypomorphic allele. Nonetheless, this variant has also been reported in trans with known pathogenic TYR variants in unaffected family members of AR OCA pedigrees and the compound heterozygote status of this patient for the TYR gene is responsible for his clinical manifestation in our patient.

PrgmNr 3444 - Sengers syndrome: Novel pathogenic variant in the AGK gene and analysis of phenotypic variability and clinical outcomes

View session detail

Author Block: **M. Caha**[1], C-H. W. Wu[2], L. B. Smoot[3], D. J. Harris[2], A. E. Roberts[3,2], S. Sacharow[2], O. A. Bodamer[2]; [1]Dept. of Pediatrics, SUNY Downstate Hlth.Sci. Univ., Brooklyn, NY, [2]Div. of Genetics and Genomics, Boston Childrenâ□□s Hosp./Harvard Med. Sch., Boston, MA, [3]Dept. of Cardiology, Boston Childrenâ□□s Hosp./Harvard Med. Sch., Boston, MA
Disclosure Block: **M. Caha:** None.
Sengers syndrome is a rare autosomal recessive mitochondrial disease caused by biallelic pathogenic variants in the AGK gene, which encodes the acylglycerol kinase enzyme. The syndrome was originally defined by the triad of hypertrophic cardiomyopathy, cataracts, and lactic acidosis with or without skeletal myopathy. Here, we present a female infant with Sengers syndrome due to a novel homozygous frameshift variant (AGK: c.1141_1142dupGG, p.S382Afs*17), who died at five months of age. Next, we provide a systematic review of the molecularly confirmed cases. We demonstrate variable expressivity and penetrance of the central features of Sengers syndrome as follows â□□ cataracts: 97.30 %, cardiomyopathy: 86.49 %, lactic acidosis: 83.78 %, skeletal myopathy: 62.16 %. Finally, we analyze associations between genotype, biochemical findings, and clinical outcomes, focusing on infantile death. Patients with homozygous nonsense variants had a tendency towards infantile death and lower age of death ($p = 0.02$ and $p = 0.06$). The location of pathogenic variants within AGK domains was not significantly associated with infantile death ($p = 0.52$). However, biochemically, absence of lactic acidosis was significantly ($p = 0.02$) associated with longer survival. Thus, lactic acidosis may potentially serve as a negative prognostic biomarker for Sengers syndrome.

PrgmNr 3445 - Whole genome sequencing of mitochondria: A powerful tool for clinical diagnosis of human mitochondria diseases

View session detail

Author Block: **J. Yang**[1], Y. Jiang[1], L. Wang[1], L. Lu[1], J. Xiao[1], C-Y. Kao[1], E. S. Schmitt[1], L. Li[1], F. Xia[2,1], L. Meng[2,1], H. Dai[2,1], S. G. Peacock[1], G. Wang[1], J. Lattier[1], J. Zhang[2,1], Z. Chen[2,1], W. J. Craigen[2,1], L-J. C. Wong[3,1], C. M. Eng[2,1], Y. Wang[2,1]; [1]Baylor Genetics, Houston, TX, [2]Baylor Coll. Med., Houston, TX, [3]Baylor Coll. of Med., Houston, TX

Disclosure Block: **J. Yang:** Salary/Employment; Baylor Genetics.

Background: Mitochondria are the organelles that, amongst other functions, generate energy for cells. In addition to the human nuclear genome, the mitochondrial genome (mtDNA) is a 16569bp double helix loop that encodes 37 genes. mtDNA is inherited from the mother and mitochondria DNA dysfunction can cause a number of human diseases including mitochondrial myopathy, Maternal Diabetes and deafness (MDD), Leber's hereditary optic neuropathy (LHON), Leigh syndrome and many others.Methods: At Baylor Genetics we have established a Next Generation Sequencing (NGS)-based platform to sequence the whole genome of mtDNA. Long-range PCR of mitochondrial DNA was used to enrich the mitochondrial sequence from the nuclear genome background and was subject to Next Generation Sequencing. NGS data were processed using NextGene software to analyze detected single nucleotide variants (SNV) and large deletions in the mitochondrial genome.Results: Since 2018 we have processed 1732 clinical cases for mtDNA testing, including 843 females, 884 males and 5 unspecified cases; with the patientâ□□s age ranging from prenatal fetus to 86-year-old adults. Sample types included blood, urine, bone marrow, liver, skeletal muscle, fibroblast, amniocytes, buccal swab and extracted DNA. We detected pathogenic or likely pathogenic SNV in 97 cases and large deletions of mitochondria genomic fragments in 161 cases. 10 cases were positive for both SNV mutations and large deletions. 1484 cases were negative for both. Pathogenic variant m.3243A>G in the tRNA Leucine gene is the most prevalent mutation, which was detected 22 times. It is heteroplasmic in all cases reported, ranging from 1.5% to 94.1%.Conclusion: We have developed a robust platform to detect variants ranging from SNV to several kb deletions in the mtDNA. It is a very powerful tool to provide a molecular diagnosis for mitochondrial disorders. We also have gene panel sequencing for nuclear genome-encoded genes that impact mitochondria function, an essential adjunct to diagnose mitochondria-related diseases. Based on these observations, we are currently pursuing multiple directions in order to improve mitochondrial sequencing efficiency and accuracy.

PrgmNr 3446 - *EFEMP1* rare variants cause juvenile-onset open angle glaucoma in families from the Philippines

View session detail

Author Block: **J. L. Wiggs**[1,2], R. Collantes[1,3], M. S. Delfin[3], B. Fan[1], J. M. R. Torregosa[4], C. Siguan-Bell[4], N. Florcruz[5,6], J. D. Martinez[6], B. Masna-Hidalgo[7], V. T. Guzman[7], J. Anotado-Flores[7], F. D. Levina[8], S. C. Hernandez.[9], A. A. Collantes[10], M. Sibulo[11], S. Rong[1]; [1]Harvard Med Sch, MEEI, Boston, MA, [2]Broad Inst., Boston, MA, [3]Dept. of Ophthalmology, Manila Doctors Hosp., Manila, Philippines, [4]Dept. of Ophthalmology, Cebu Inst. of Med., Cebu Velez Gen. Hosp., Cebu City, Philippines, [5]Dept. of Ophthalmology and Visual Sci., Univ. of the Philippinesâ□□Philippine Gen. Hosp., Manila, Philippines, [6]Glaucoma Service, Dept. of Ophthalmology, East Avenue Med. Ctr., Quezon City, Philippines, [7]Dept. of Ophthalmology, Western Visayas Med. Ctr., Iloilo City, Philippines, [8]Dept. of Ophthalmology, Jose R. Reyes Mem. Med. Ctr., Manila, Philippines, [9]UmeÃ¥ Univ., UmeÃ¥, Sweden, UmeÃ¥, Sweden, [10]Coll. of Med., Our Lady of Fatima Univ., Valenzuela, Philippines, [11]Faculty of Med. and Surgery, Univ. of Santo Tomas, Manila, Philippines

Disclosure Block: **J.L. Wiggs:** Consultant/Consulting Fees/Other Remuneration; Aerpio, Allergan, Editas, RegenXbio, Avellino, Maze.

Glaucoma is the leading cause of irreversible blindness world-wide and exhibits both Mendelian (childhood onset) and complex (adult onset) inheritance. Ten genes are known to cause childhood glaucoma, but in total these only account for disease in approximately 20% of patients. Because most of the childhood glaucoma genes have been identified in European Caucasian populations, to facilitate novel gene discovery we have studied affected families from various geographic regions including the Philippines. As part of this effort, we used exome sequencing to evaluate 14 Filipino juvenile open-angle glaucoma (JOAG) families and identified 3 independent families (35, 2 and 27 members) with rare *EFEMP1* variants (p.N80Y, p.R477C and p.Ter494Glnext*29) co-segregating with disease. None of the rare variants are present in gnomAD and all modify highly conserved amino acids. Affected variant carriers (N= 34) exhibited severe disease with average age of disease onset of 16 years (range 3-43) and 76% developing blindness. Common SNPs near *EFEMP1* have been associated with adult-onset glaucoma (POAG) and a low frequency *EFEMP1* variant (p.R140W; MAF 0.0008%) may contribute to disease risk in one POAG family. Interestingly, a single *EFEMP1* missense allele (p.R345W) is known to cause Malattia Leventinese (ML), an inherited retinal degeneration. *EFEMP1* is an extracellular matrix protein with ocular expression similar to Myocilin, another extracellular matrix protein known to cause JOAG through a mechanism involving protein misfolding and endoplasmic reticulum aggregation. To determine if *EFEMP1* variants exhibit similar effects, we transfected COS7 cells with vectors expressing the three novel *EFEMP1* variants as well as variants associated with POAG and ML. We showed that all three variants found in JOAG patients caused significant intracellular protein aggregation compared to wild type and also the variants associated with the other phenotypes (p.R140W and p.R345W). These results suggest that rare coding *EFEMP1* variants can cause JOAG through a mechanism involving protein aggregation and that the extent of intracellular protein aggregation and retention appears to be the basis for the observed phenotype spectrum. This is the first report of *EFEMP1* variants causing early-onset glaucoma and we show that *EFEMP1* variation appears to be a relatively common cause of childhood glaucoma in these Filipino families. These results underscore the value of ethnically diverse populations for comprehensive detection of disease-causing mutations.

PrgmNr 3447 - *FOXP1* haploinsufficiency contributes to the development of congenital diaphragmatic hernia

View session detail

Author Block: **K. Pendleton**[1,2], A. Hernandez-Garcia[1], J. M. Lyu[3,4], I. M. Campbell[5], C. Shaw[6], F. A. High[4,3,7], P. K. Donahoe[3,8], W. K. Chung[9,10], D. A. Scott[1,11]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Genetics and Genomics Graduate Program, Baylor Coll. of Med., Houston, TX, [3]Pediatric Surgical Res. Lab., Massachusetts Gen. Hosp., Boston, MA, [4]Dept. of Surgery, Boston Children's Hosp., Boston, MA, [5]Div. of Human Genetics, Children's Hosp. of Philadelphia, Philadelphia, PA, [6]Baylor Coll. of Med., Houston, TX, [7]Dept. of Pediatrics, Massachusetts Gen. Hosp., Boston, MA, [8]Dept. of Surgery, Harvard Med. Sch., Boston, MA, [9]Dept. of Pediatrics, Columbia Univ., New York, NY, [10]Dept. of Med., Columbia Univ., New York, NY, [11]Dept. of Molecular Physiology and Biophysics, Baylor Coll. of Med., Houston, TX

Disclosure Block: **K. Pendleton:** None.

Congenital diaphragmatic hernia (CDH) is a common birth defect that is often accompanied by life-threatening lung hypoplasia leading to respiratory complications. A variety of chromosomal abnormalities, genomic alterations, and sequence variants have been shown to cause or predispose to the development of CDH. Despite these discoveries, the genetic factors underling most cases of CDH remain unidentified. The forkhead box 1 gene (*FOXP1*) encodes a transcriptional repressor involved in tissue regulation and cell type-specific functions during development. Heterozygous *FOXP1* variants have been shown to cause a variety of structural birth defects including central nervous system anomalies, congenital heart defects, congenital anomalies of kidney and urinary tract, cryptorchidism, and hypospadias. Using data from individuals with CDH who carry interstitial deletions, we have defined a CDH critical region on chromosome 3p13 that contains five protein coding genes. Of these genes, only *FOXP1* is 1) expressed in the developing diaphragm, 2) exhibits high loss-of-function intolerance, 3) has a high level of similarity to known CDH genes as defined by a previously published machine learning algorithm. We have also identified two individuals with CDH who carry de novo, pathogenic variants in *FOXP1* that are predicted to trigger nonsense mediated mRNA decay. Since CDH is only seen in a subset of individuals with *FOXP1* haploinsufficiency, it is likely that loss of *FOXP1* functions as a susceptibility factor that contributes to the development of CDH in conjunction with other genetic, epigenetic, environmental, and/or stochastic factors. In conclusion, we define a CDH critical region on chromosome 3p13 and provide evidence that *FOXP1* haploinsufficiency can contribute to the development of CDH in humans.

PrgmNr 3448 - A multidisciplinary nephrogenetic referral clinic for children and adults - diagnostic achievements and insights

View session detail

Author Block: **B. Pode-Shakked**[1,2], Y. Ben-Moshe[1,2], O. Barel[3], L. C. Regev[1,2], M. Kagan[1,2], A. Eliyahu[3,2], D. Marek-Yagel[1,2], D. Atias-Varon[1], E. Lahav[1,2], N. Issler[1,2], O. Shlomovitz[1,2], P. Beckerman[3,2], M. Greenberg[4,2], O. Volovelsky[5], Y. Borovitz[6,2], M. Davidovits[6,2], O. Haskin[6,2], H. Alfandary[6,2], S. Levi[6,2], M. Kaidar[7,2], Z. Katzir[8,9], A. Angel-Korman[8,9], A. Leiba[8,9], D. Lotan[1,2], M. Shashar[10,11], Y. Anikster[12,2], A. Raas-Rothschild[1,2], G. Rechavi[3,2], B. Dekel[1,2], A. Vivante[1,2]; [1]Edmond and Lily Safra Children's Hosp., Sheba Med. Ctr., Ramat Gan, Israel, [2]Sackler Faculty of Med., Tel-Aviv Univ., Tel-Aviv, Israel, [3]Sheba Med. Ctr., Ramat Gan, Israel, [4]Meir Med. Ctr., Kfar Saba, Israel, [5]Hadassah Med. Ctr. and Faculty of Med., Hebrew Univ. of Jerusalem, Jerusalem, Israel, [6]Schneider Children's Med. Ctr. of Israel, Petach Tikva, Israel, [7]Edith Wolfson Med. Ctr., Holon, Israel, [8]Samson Assuta Univ. Hosp., Ashdod, Israel, [9]Ben-Gurion Univ., Beer Sheva, Israel, [10]Laniado Hosp., Netanya, Israel, [11]Ruth and Bruce Rappoport Faculty of Med., Technion Inst., Haifa, Israel, [12]Edmond and Lily Safra Children's Hosp., Tel Hashomer, Israel
Disclosure Block: **B. Pode-Shakked:** None.
BACKGROUND Genetic kidney diseases contribute a significant portion of kidney diseases in children and young adults. Nephrogenetics is a rapidly evolving subspecialty, however, in the clinical setting, increased use of genetic testing poses implementation challenges. Consequently, we established a national nephrogenetic clinic at Sheba Medical Center to apply a multidisciplinary model. **METHODS** Patients were referred for clinic evaluation from different pediatric or adult nephrology units across the country, if their primary nephrologist suspected an undiagnosed genetic kidney disease. We determined the diagnosis rate among patients who were referred to the clinic and observed the effect of diagnosis on medical care. We also discuss the requirements of a nephrongenetic clinic in terms of logistics, recommended indications for referral and building a multidisciplinary team. **RESULTS** Over a period of 24 months, a total of 103 probands were evaluated at the clinic, with an age range of 10 days to 72 years. The most common phenotypes included congenital anomalies of the kidneys and urinary tract, nephrotic syndrome or unexplained proteinuria, nephrocalcinosis/nephrolithiasis, tubulopathies and cryptogenic end-stage kidney disease. Over 90% of accepted patients were referred due to clinical suspicion of an undetermined underlying genetic diagnosis. A molecular diagnosis was reached in 34/53 probands for whom genetic evaluation was completed, yielding a diagnostic rate of 64.2%. Of these, over 65% of the diagnoses were made via next generation sequencing (gene panel- or exome sequencing). **CONCLUSIONS** We identified a significant fraction of genetics kidney etiologies among previously undiagnosed individuals which influenced subsequent clinical management. Our results support that nephrogenetics, a rapidly evolving field, may benefit from well-defined multidisciplinary co-management administered by a designated team of nephrologist, geneticist and bioinformatician.

PrgmNr 3449 - Carrier Prevalence of 302 genetic diseases in the Mexican Jewish Community

View session detail

Author Block: **D. Morgenstern-Kaplan**[1], J. Raijman-Policar[1], S. Majzner-Aronovich[1], S. Aradhya[2], D. Pineda-Alvarez[2], E. GarcÃa-Vences[1], M. Aguinaga[3,4]; [1]Centro de InvestigaciÃ³n en Ciencias de la Salud AnÃihuac (CICSA),Univ. AnÃihuac MÃ©xico, Huixquilucan, Mexico, [2]Invitae, San Francisco, CA, [3]Centro de InvestigaciÃ³n en Ciencias de la Salud AnÃihuac (CICSA),Univ. AnÃihuac MÃ©xico, Huixquilucan, Estado de Mexico, Mexico, [4]Dept.o de Salud Sexual y ReproducciÃ³n Humana, Inst. Natl. de PerinatologÃa, Mexico City, Mexico
Disclosure Block: **D. Morgenstern-Kaplan:** None.
Background: The Mexican Jewish community (MJC) is a previously uncharacterized genetically isolated group composed of Ashkenazi, Sephardi and Mizrahi Jews that migrated from different countries in the early 1900s. Historically, these populations have been considered to be at high risk of being carriers of rare genetic diseases. Preconception carrier screening is recommended by the American College of Obstetrics and Gynecology and the American College of Medical Genetics and Genomics to identify carriers and couples who are at risk of having an affected child. The carrier frequencies of most genetic diseases in the MJC and in most Latin-American Jewish communities is unknown.
Methods: We conducted a cross-sectional study of 208 MJC individuals between ages 18-40 years old representing Ashkenazi (AJ), Sephardi-Mizrahi (SMJ), or mixed ancestry (MAJ) origin. We offered saliva-based preconception pan-ethnic carrier screening that examined 302 genes using next-generation sequencing (NGS) to determine the prevalence of carriers. The study protocol was approved by an IRB and a bioethics committee at our institution, and informed consent was obtained from all participants.
Results: The carrier screening results showed that 86.5% of participants had at least one pathogenic/likely pathogenic (P/LP) variant in one of the 302 genes analyzed, with some participants being carriers for as many as seven different genetic diseases. The most common genes with pathogenic variants were HFE (23.5%), CFTR (16.8%), MEFV (11.5%), F5 (10%) and GALT (10%). Furthermore, when analyzed in subgroups, the most common genes with P/LP variants in AJ were HFE (33.7%), WNT10A (11.6%) and CFTR (10.3%). The most common genes with P/LP variants in SMJ were HFE (19.6%), F5(18.1%) and CFTR (16.6%). The most common genes with P/LP variants in MAJ were CFTR (24.6%), MEFV (16.9%) and HFE (15.3%). Finally, 18.2% of couples were identified to be at risk of having an affected child due to a commonly shared variant (73.3% of at-risk couples) or due to two separate P/LP variants in the same gene (26.6% of at-risk couples).
Conclusions: The prevalence of carriers of at least one genetic disease in the MJC was 86.5%, and 18.2% of couples detected were at risk of having an affected child. The study of this population, along with other Latin-American Jewish populations using an NGS based carrier screening panel will help future parents to make more informed reproductive decisions.

PrgmNr 3450 - Congenital diarrhea is a common cause of neonatal crisis and up to ~60% of cases among neonates may be inherited

View session detail

Author Block: **J. Tumolo**, H. Druker, J. L. Weber; Prevention Genetics, Marshfield, WI
Disclosure Block: **J. Tumolo:** Salary/Employment; Prevention Genetics.
Chronic diarrhea is a symptom in a group of clinically and genetically heterogeneous disorders. Studies suggest approximately 7 in 1,000 newborns admitted to neonatal intensive care units are admitted for diarrhea. Of these, approximately 60% of cases are acquired, ~15% are inherited, and the remainder result from an unknown etiology (Passariello et al. 2010. PubMed ID: 20518089). Hereditary congenital diarrhea and enteropathies are characterized by persistent, severe diarrhea that often presents within the first few months of life, is life-threatening, and requires medical intervention in the form of parenteral (IV) nutrition or hydration. At the molecular level, these disorders result from disruption of intestinal epithelial cell function or from immune system dysfunction that secondarily impairs the intestinal epithelium (Canani et al. 2015. PubMed ID: 25782092). While intractable diarrhea is the major clinical feature, congenital diarrheas and enteropathies can have extraintestinal manifestations as well as systemic disease. The early identification of a genetic cause of congenital diarrhea or enteropathy helps to inform a treatment strategy such as hematopoietic stem cell transplantation, enzyme replacement therapy, a special diet or other pharmaceutical support. Forty-one patients (age 2 weeks old to 62 years) with severe diarrhea underwent genetic testing using a 114 gene next generation sequencing (NGS) panel performed in a CLIA approved clinical laboratory. Of the patients tested, 24% were â¤ 30 days old, 51% were CFTR, suggesting these patients had a previously unknown diagnosis of cystic fibrosis. Among definite or potential diagnostic findings, 35% had a disorder related to immune system dysfunction, while 65% were related to disordered epithelial cell function. Importantly, 64% of definite or potential molecular diagnoses were medically actionable. In summary, genetic testing of individuals with severe diarrhea led to a definite or potential diagnosis in 24% of patients but this number increased to 60% in patients presenting in the neonatal period. In 64% of diagnosed patients, the results informed patient care, supporting the clinical utility of early genetic testing in this patient cohort.

PrgmNr 3451 - Diagnostic rate of clinical exome sequencing for CAKUT phenotypes

View session detail

Author Block: **E. Rivera Munoz**[1], J. A. Rosenfeld[1,2], J. E. Posey[1], D. A. Scott[3]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Genetics, Houston, TX, [3]Baylor Coll. Med, Houston, TX
Disclosure Block: **E. Rivera Munoz:** None.

Congenital Anomalies of Kidney and Urinary Tract (CAKUT) are estimated to make up 40-50% of pediatric malformations and are a predominant cause of kidney failure in children and young adults. This category of malformations describes a broad range of phenotypes including renal agenesis, duplication of the collecting system, multicystic dysplastic kidney, ureteropelvic junction obstruction, vesicoureteral reflux, and others. CAKUT results from aberrant embryonic development and can be induced by both genetic and environmental factors. Familial studies show that putative pathogenic variants often demonstrate incomplete penetrance and variable expressivity. Although exome sequencing is commonly used to identify a molecular cause in individuals with congenital anomalies, this test is not universally ordered in individuals with CAKUT. Many studies have shown positive results using natural language processors (NLP) to extract phenotypic information from clinical notes. Additional valuable insight may be gained by using these tools that would otherwise require highly labor-intensive processes. Machine learning tools have also been used to harness annotations from knowledgebases to help identify phenotypic expansions in clinically relevant genes. These additional layers of evidence may inform variant interpretation and improve the diagnosis and prognosis of CAKUT.

In the present study, we retrospectively analyzed molecular findings in individuals who were referred for clinical exome sequencing that included one or more CAKUT features. This clinical laboratory dataset contains age, sex, clinical features (as free-text descriptors), and the set of variants that were reported back to referring clinicians as potentially explaining part or all clinical features. Using Doc2HPO, we extracted phenotype terms from free-text and used them to categorize these cases by renal features. From the entire clinical laboratory cohort of ~15,000 individuals, we identified 925 individuals with one or more reported CAKUT features with 95% being non-isolated. The most prevalent features in our cohort were: >1 CAKUT feature (38%), renal hypoplasia/aplasia (12%), renal cyst (11%), abnormal kidney localization (9%), renal insufficiency (6%), abnormal renal pelvis morphology (5%), and nephrolithiasis (3%). Of these cases, 195 (21%) were considered solved with an additional 243 (26%) being classified as possibly or partially solved. The top returned genes included *PKD1*, *PKHD1*, *KMT2D*, *TMEM231,* and *CC2D2A*. Based on these results, we conclude that clinical exome sequencing is an effective molecular diagnostic test for individuals with non-isolated CAKUT.

PrgmNr 3452 - Exome sequencing further implicates disruption of the *NODAL* signaling pathway in laterality defects

View session detail

Author Block: **M. Dawood**[1,2,3], S. Bowne[4], J. M. Fatih[1], A. Jolly[1,3], H. Du[1], Z. Dardas[1], C. M. Grochowski[1], S. N. Jhangiani[1,2], X. H. T. Wehrens[5,6,7,8], N. Hanchard[1], P. Liu[1], E. Boerwinkle[2,4], J. Posey[1], D. M. Muzny[1,2], R. A. Gibbs[1,2], J. R. Lupski[1,2,9,10], S. A. Morris[8], Z. Coban Akdemir[1,4]; [1]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [2]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [3]Med. Scientist Training Program, Baylor Coll. of Med., Houston, TX, [4]Human Genetics Ctr., Dept. of Epidemiology, Human Genetics, and Environmental Sci., Sch. of Publ. Hlth., The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [5]Cardiovascular Res. Inst., Baylor Coll. of Med., Houston, TX, [6]Dept. of Molecular Physiology and Biophysics, Baylor Coll. of Med., Houston, TX, [7]Dept. of Med., Baylor Coll. of Med., Houston, TX, [8]Dept. of Pediatrics, Div. of Cardiology, Texas Children's Hosp. and Baylor Coll. of Med., Houston, TX, [9]Texas Children's Hosp., Houston, TX, [10]Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX

Disclosure Block: **M. Dawood:** None.

Heterotaxy is a genetically and phenotypically heterogeneous laterality defect characterized by abnormal left-right patterning with the most common, clinically affected organ systems being the cardiovascular system, gastrointestinal tract, spleen, and liver. Pathogenic missense mutations in *NODAL* have been previously implicated as causative for autosomal dominant heterotaxy (MIM: 270100). Null mutations in *Nodal* arrest mouse embryonic development at the gastrulation stage, but mice with conditional hypomorphic *Nodal* mutants show phenotypes consistent with a perturbed embryonic left-right axis. Here, we analyzed a cohort of 322 proband-only exomes of individuals with clinically diagnosed laterality defects using family-based, rare variant analyses. We found evidence for missense, frameshift, and/or structural variant perturbations in the *NODAL* gene as potential causes of heterotaxy and other laterality defects in 27 unrelated probands. Our data from enrichment analyses of this cohort implicate likely damaging variants in members of the *NODAL* signaling pathway and the broader TGF-Î² superfamily. Further, we reinvestigated the *NODAL* missense mutation NM_018055.5:c.778G>A:p.G260R which was originally described as causative of heterotaxy with reduced penetrance, variable expressivity, and predominantly affecting Hispanic individuals. Mohapatra *et al.* in 2009 first implicated G260R as pathogenic on the basis of three lines of functional evidence, but since then G260R has been reclassified to â□□conflicting interpretations of pathogenicityâ□□ based on the American College of Medical Genetics and Genomics criteria (ClinVar Accession: VCV000008269.9). We describe the underlying basis of the consternation and provide further confirmatory evidence in support of pathogenicity. In aggregate, our study uncovers evidence supporting the prominence of the *NODAL* pathway in embryological left/right axis pattern perturbations and birth defects.

PrgmNr 3453 - Identifying susceptibility genes for congenital diaphragmatic hernia (CDH) using a machine learning algorithm

View session detail

Author Block: **A. Hardcastle**[1], I. M. Campbell[2], V. Varghese[3], N. L. Shannon[4], S. A. McKee[5], P. D. Turnpenny[6], J. Morton[7], J. Eason[8], C. A. Shaw[9], D. A. Scott[9]; [1]Brigham Young Univ., Provo, UT, [2]Children's Hosp. of Philadelphia, Philadelphia, PA, [3]All Wales Med. Genomics Service, Cardiff, United Kingdom, [4]Nottingham City Hosp., Nottingham, United Kingdom, [5]Belfast City Hosp., Belfast, United Kingdom, [6]Royal Devon & Exeter Hosp, Exeter, United Kingdom, [7]Birmingham Women's Hosp., Birmingham, United Kingdom, [8]City Hosp. Campus, Nottingham, United Kingdom, [9]Baylor Coll. of Med., Houston, TX

Disclosure Block: **A. Hardcastle:** None.

Congenital diaphragmatic hernia (CDH) is a common, life-threatening birth defect seen in approximately 1 in 4000 newborns. A molecular etiology can only be identified in a subset of CDH cases. To identify new CDH susceptibility genes, we prioritized candidate genes identified through a search of the DECIPHER database using a machine learning algorithm that integrates data from Gene Ontology (GO), the Mouse Genome Database (MGI), the Protein Interaction Network Analysis (PINA) platform, the Kyoto Encyclopedia of Genes and Genomics (KEGG) molecular interaction network, published miRNA targets, the GeneAtlas expression distribution, and the NIH Roadmap Epigenomics Mapping Consortium. In our search we identified individuals with CDH who carried putatively deleterious sequence variants in *FREM2*, *SMARCA4*, *CREBBP*. All of these genes were shown to have high similarity to known CDH genes using our machine learning algorithm (99.8, 92.2, and 99.7th centile compared to all RefSeq genes, respectively). The role of these genes in diaphragm development is further supported by their expression in the developing mouse diaphragm, by mouse models of CDH development, and/or previously published case reports. Our results suggest that our machine-learning algorithm can be used to identify CDH susceptibility genes, and that loss of *FREM2*, *SMARCA4*, and *CREBBP* function each leads to an increased risk of developing CDH.

PrgmNr 3454 - Increased Expression of ZFPM2 Bypasses SRY to Drive 46,XX Testicular Development: A New Mechanism of 46,XX DSD

View session detail

Author Block: **L. Ragno**[1], D. Yang[1], T. R. Bhatti[2], J. P. Bradfield[2,3], D. K. Bowen[4], K. L. Grand[2], H. Hakonarson[2], T. F. Kolon[4], J. Glessner[2], M. C. Simon[5], N. Skuli[5], M. G. Vogiatzi[2], M. A. Guerraty[6], M. A. Deardorff[2], L. C. Pyle[2]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Div. of Human Genetics, Dept. of Pediatrics, Children□□s Hosp. of Philadelphia, Philadelphia, PA, [3]Quantinuum Res., LLC, Wayne, PA, [4]Div. of Pediatric Urology, Dept. of Urology, Children□□s Hosp. of Philadelphia, Philadelphia, PA, [5]Dept. of Cell and Dev.al Biology, Univ. of Pennsylvania, Philadelphia, PA, [6]Div. of Cardiovascular Med., Dept. of Med., Philadelphia, PA

Disclosure Block: **L. Ragno:** None.

We present a patient with a novel cause of 46,XX ambiguous/androgenous genitalia Differences of Sex Development (DSD). Genome-wide array from blood showed 46,XX with ~35% mosaic duplication of 76.5Mb at chromosome 8q13.2-q24.3, containing 257 OMIM genes including *ZFPM2* and *CYP11B1*. Congenital adrenal hyperplasia testing was negative, testosterone was elevated, and the pro-testicular master regulator *SRY* was absent. The infant had a uterus, one streak ovary (We hypothesized that mosaic ZFPM2 duplication resulted in localized ZFPM2 overexpression and testicular development. In typical testicular development, *ZFPM2* and its binding partner, *GATA4*, drive expression of the *SRY* master regulator. We completed RNA-seq of twelve single-cell clonal fibroblast cell lines from this mosaic individual to identify differentially regulated genes (DRGs). After quality control, three lines representing with and without the duplication were analyzed. Nine (9; 3.5%) genes within the duplicated gene were significantly overexpressed >1.5-fold, versus 78 genes across the rest of the genome (0.47%) (significance pSRY or *CYP11B1* (CAH), as expected. *ZFPM2* was the second most highly overexpressed gene (FC=94.7, p=3.56x10-57). Thirty (30) genes were significantly under-expressed (RSPO1 (FC=-2.26, p=6.0x10-3), and the pro-testicular transcription factor *MAP3K1* was upregulated (FC=2.07, p=3.54x10-5). Yet further genes involved in testicular development were upregulated, including *AKR1C2* (chr10), an enzyme of non-canonical dihydrotestosterone synthesis (FC=2.52, p=7.4x10-3). We have shown that increased *ZPFM2* dosage can induce 46,XX testicular development in a manner not dependent on *SRY*. This contravenes the previous understanding that GATA4/ZFPM2 drives testicular development through *SRY*. *ZFPM2* may modulate numerous critical sex development genes including transcription factors otherwise thought to be downstream of *SRY* (*MAP3K1*, *AKR1C2*). Findings from this single high-yield patient demonstrate that the primary role of *ZFPM2* in testicular development may be independent of *SRY*. This adds *ZFPM2* to the brief (SRY. Overall, new understanding of these genes demonstrates that the role of *SRY* as a â□□master regulator' of testicular development may be less than previously thought.

PrgmNr 3455 - Patient with novel compound heterozygous variants in *NCKAP1L* presents Immunodeficiency 72 with autoinflammation and novel features

View session detail

Author Block: **G. F. Godinez-Zamora**[1,2], E. VÃisquez-Echeverri[3], O. Saucedo-RamÃrez[4], P. Baeza-Capetillo[2,5], V. MorÃin-Barroso[6], J. Aguirre-HernÃindez[2]; [1]Univ. Natl. Autonoma de MÃ©xico, Mexico city, Mexico, [2]Laboratoy of Genomics, Genetics and Bioinformatics, Hosp. Infantil de MÃ©xico, Mexico city, Mexico, [3]Lab. of Immune Deficiencies, Inst. Natl. de PediatrÃa, Mexico city, Mexico, [4]Allergy and Immunology Dept., Hosp. Infantil de MÃ©xico, MÃ©xico, Mexico city, Mexico, [5]Dept. of Genetics, Hosp. Infantil de MÃ©xico, MÃ©xico, Mexico city, Mexico, [6]Hosp. Gen. de MÃ©xico Dr. Eduardo Liceaga, Mexico city, Mexico

Disclosure Block: **G.F. Godinez-Zamora:** None.

Nck-associated protein 1-like (*NCKAP1L*) encodes a protein that regulate actin cytoskeleton in hematopoietic cells. Last year, biallelic alterations in this gene were identified as the cause of immunodeficiency 72 with auto inflammation (OMIM 618982), which is characterized by immunodeficiency, inflammation and lymphoproliferation with features of hemophagocytic lymphohistiocytosis (HLH), and atopic disease. Other clinical features include recurrent infections, elevated IgE levels, hepatosplenomegaly, lymphadenopathy, bronchiectasis, and autoimmunity; which were present in known patients since their first years of life. Here we present a patient with two variants in *NCKAP1L* (one of them novel) and some phenotypic characteristics that sets her apart from the 9 patients known in the literature. Our patient is an adolescent female, with no apparent health issues excepting acute pyelonephritis at the age of 4 years. At 14 years, she was diagnosed with systemic lupus erythematosus due to malar rash, arthritis, grade IV proliferative lupus nephritis, high ANA concentration; and Anti-Sm. She present arterial hypertension; difficult to manage, and transient brain ischemic events associated with seizures, which may be secondary to inflammation of the central nervous system. Laboratory workup reported: leucopenia, bandemia, lymphopenia and trombocytopenia. Low levels for C3, C4, hypogammaglobulinemia (IgG, IgM, IgA) and normal IgE. In contrast to the patients described in the literature, she had normal IgE levels, normal hepatic function, and no hepatosplenomegaly. Moreover, she hasnâ□□t presented lymphoproliferation, HLH or lymphadenopathy. Whole exome sequencing was performed. Variants were filtered by depth, allelic fraction, consequence, frequency in public databases, and in a local database. Variants were searched in all known primary immunodeficiency genes, and a compound heterozygous genotype, with two missense variants, was found in *NCKAP1L*:c.1492G>A:p.Val498Met, c.3346G>A:p.Ala1116Thr. The first is a known variant with a very low frequency in gnomAD (4 times in v3.1.1). The second variant is not found in any public databases. A very small number of patients have been diagnosed with immunodeficiency 72. Most share a common set of phenotypic characteristics, and this may suggest that the phenotype may be rather uniform. However, we have identified an adolescent female, with two variants in *NCKAP1L*, who doesnÂ´t have some of the phenotypic alterations observed in the rest of the patients with this disease. These may suggest that the phenotypic spectrum of this immunodeficiency may be broader than initially thought.

PrgmNr 3456 - Precision medicine models for undiagnosed and rare disease

View session detail

Author Block: **L. C. Burrage**[1], J. A. Rosenfeld[2], S. Yamamoto[1], M. F. Wangler[2], M. Roth[2], D. Lanza[2], R. Marom[1], B. Lee[2], A. Milosavljevic[1], J. Heaney[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. Med., Houston, TX

Disclosure Block: **L.C. Burrage:** None.

Exome, genome and RNA sequencing have transformed our ability to diagnose patients with suspected genetic disease. These technologies have led to the discovery of hundreds of new disease genes and to phenotypic expansion within known diagnoses. However, up to 70% of individuals with suspected genetic disease remain undiagnosed likely because their disease-causing variant(s) has yet to be discovered or the clinical significance of variants is unclear. The Undiagnosed Diseases Network (UDN) Model Organism Screening Center (MOSC) has demonstrated that various genome modification techniques in model organisms (*C. elegans*, *Drosophila*, *D. rerio*) are important tools aiding in the interpretation of variants of uncertain significance. For example, fly models established the pathogenicity and functional consequences of variants in *CACNA1A* (R1673P and R1664Q) associated with early onset cerebellar ataxia and developmental delay. Studies in the mouse offer a complementary approach to these other model organisms and open new avenues to test therapy and disease mechanism in a mammalian model. Critically, a variety of existing resources from the Knockout Mouse Phenotyping Program (KOMP2) have supported gene discoveries. For example, existing data generated by KOMP2 for *Otud6b*$^{-/-}$ mice supported the pathogenicity of biallelic loss-of-function variants in *OTUD6B*. The finding of perinatal lethality, small size, and ventricular septal defects in the *Otud6b*$^{-/-}$ mice is consistent with the human phenotype, which includes failure to thrive and cardiac defects. Likewise, low bone density in *Copb2*$^{+/-}$ mice supported the pathogenicity of heterozygous loss-of-function *COPB2* variants in individuals with low bone density. Lastly, early lethality and failure to progress to blastocyst stage in *Tonsl*$^{-/-}$ embryos supported the pathogenicity of biallelic hypomorphic variants in this essential gene in individuals with a skeletal dysplasia. Given that fly models can rapidly assess clinical significance of a variant and mouse models can provide insight into variant disease mechanisms and assess treatment strategies in the context of mammalian biology, we established the BCM Center for Precision Medicine Models to capture the benefits associated with each model system and the synergy derived from having two distinct yet complementary functional readouts. The center is working with local, national, and international programs and researchers on the development of precision models that will end the diagnostic odyssey of patients with undiagnosed, rare, and Mendelian diseases and launch pre-clinical studies investigating personalized medicine strategies.

PrgmNr 3457 - Rapid CRISPR/Cas9-based method for functional validation of human disease genes

View session detail

Author Block: **S-J. Lin**, C. Petree, K. Huang, G. K. Varshney; Oklahoma Med. Res. Fndn., Oklahoma City, OK
Disclosure Block: **S. Lin:** None.
Since the completion of the Human Genome Project and the with the decreasing cost of whole-genome sequencing technologies, thousands of whole-exome (WES) and whole-genome sequencing (WGS) projects have advanced our understanding of the causes of genetic disorders as they identify the full spectrum of mutations associated with human disease genes. Human geneticists now face the immense challenge of validating these candidate disease genes. Identifying which candidate genes or loci are pathologically relevant represents a critical barrier that limits our ability to generate relevant human disease models. Model organisms have revolutionized our understanding of the human disease; inactivation of candidate human disease genes in animals (i.e. creating gene â□□knockoutsâ□□) often triggers analogous phenotypes and provides a valuable disease model. CRISPR/Cas9 based methods are being used to screen for phenotypes in F0 (the founding generation) by generating biallelic mutations; we further optimized this strategy to maximize phenotype penetrance and use this method in a high-throughput manner. We combined the mutagenesis and tissue-specific phenotyping approach using transgenic reporter lines to functionally test the candidate disease genes, and identified genes involved in neurodevelopmental, musculoskeletal, vasculature, hearing, and digestive disorders. This strategy has enabled us to screen for hundreds of candidate disease genes rapidly with modest resource investment.

PrgmNr 3458 - Refined SNP haplotype analysis uncovers a potential founder deletion mutation in the *Cerebral Cavernous Malformations* 2 gene

View session detail

Author Block: **C. Gallione**[1], M. R. Detter[2], C. Lee[3], D. A. Marchuk[1]; [1]Duke Univ Med Ctr, Durham, NC, [2]Duke Univ Med Ctr., Durham, NC, [3]Angioma Alliance, Charlottesville, VA
Disclosure Block: **C. Gallione:** None.
Cerebral cavernous malformations (CCM) are vascular malformations consisting of collections of enlarged capillaries occurring in the brain or spinal cord. These vascular malformations can occur sporadically or susceptibility to develop these can be inherited as an autosomal dominant trait due to heterozygous mutation in one of three genes. Over a decade ago, we described a large germline 77.6kb deletion spanning exons 2-10 in the *CCM2* gene found in multiple affected individuals from seemingly unrelated families. Haplotype analysis using linked, microsatellite markers showed that this deletion was present on a common microsatellite haplotype but the deletion may have arisen at least twice independently. In the ensuing decades, many more CCM patients have been identified with this same deletion and in this present study, we examined genomic DNA from 25 reportedly unrelated affected individuals with this deletion. To investigate the origin and relationship of the deletion at base pair level resolution, we sequenced approximately 10 kb upstream and downstream from the recombination junction on the deleted (affected) alleles. All patients showed the identical SNP haplotype across this combined 20 kb interval. Intriguingly, all samples harbor the minor allele at rs7792895 (T=10%, GnomAD) near the deletion, making it even more likely that the deletions have a common origin. In parallel, genealogical records have traced 10 individuals in this current cohort back to five separate pedigrees dating as far back as the 1600-1700â□□s. Based on this new SNP haplotype data and our genealogical investigation, it is likely that these families and the remaining â□□unrelatedâ□□ samples converge on a common ancestor due to a founder mutation occurring hundreds of years ago on the North American continent.

PrgmNr 3459 - Significant burden of *de novo* damaging variants in novel genes in patients with congenital kidney malformations

View session detail

Author Block: **H. Milo Rasouly**[1], S. Krishna Murthy[2], S. Bedha[1], D. Fasel[3], M. Marasa[1], E. Fiaccadori[4], A. Materna-Kiryluk[5], G. Masnata[6], V. Tasic[7], M. Saraga[8], K. Kiryluk[3], G. Ghiggeri[9], S. Sanna-Cherchi[10], A. G. Gharavi[10]; [1]Columbia Univ., New York, NY, [2]Irving Med. Ctr. , Columbia Univ., New York,, NY, [3]New York, NY, [4]Univ. degli Studi di Parma Dipartimento di Med. e Chirurgia, Emilia-Romagna, Parma, Italy, [5]Poznan Univ. of Med. Sci., Poznan, Poland, [6]Azienda Ospedaliera Brotzu, Cagliari,, Italy, [7]Univ. Children's Hosp., Skopje, Macedonia, The Former Yugoslav Republic of, [8]Univ. of Split, Split, Croatia, [9]Istituto Giannina Gaslini, Genova,, Italy, [10]Columbia Univ, New York, NY
Disclosure Block: **H. Milo Rasouly:** None.
Background: Congenital kidney malformations (CKM) are one of the most common cause of pediatric kidney failure. While multiple causative genes have been identified, they explain only 10-15% of cases. *De novo* variantsâ (DNV) analysis led to the identification of novel genes for congenital heart defects and neurodevelopmental disorders. We hypothesized that similar analyses could identify new CKM-causing genes. Methods: Patients enrolled through our study on genetics of kidney diseases underwent Exome sequencing (ES) or Genome sequencing (GS). ES was performed on 151 trios (i.e. affected child with CKM and unaffected parent), and GS on 53 additional trios.100 CKM trios were identified from the Deciphering Developmental Delay consortia (ES). The sequences were processed using a BWA/GATK 4.1 pipeline on the CAVATICA platform and annotated through Ensembl VEP. Custom bioinformatics pipelines were then used for data cleaning. Potential enrichment for DNV was analyzed with the denovolyzer package in R. Results: Out of the 304 trios analyzed, 10 cases (3%) had a DNV in one of the 172 genes known to be associated with dominant forms of kidney disease (3.2 fold-enrichment, p-value=1.41 x 10^{-3}). Those 10 cases included one protein-truncating variant (PTV) in *PAX2*, one in *HNF1B* and two PTV DNVs in *KAT6B*. Genome-wide, we observed a significant 2.8 fold-enrichment for PTV DNV (76 DNVs compared to 26.6 expected, p-value=4.3x10^{-15}) and a significant 1.3 fold-enrichment for missenses DNV (248 DNVs compared to 191.5 expected, p-value=5.1x10^{-5}). When constraining the analysis to genes highly expressed in nephron-progenitor cells at 18 weeks of gestation (Human fetus) and not known to be associated with kidney disease in Human, we observed significantly increased enrichment for PTV DNV (8 DNVs compared to 1.2 expected, OR=6.9, p=3.04 x 10^{-5}, list of 593 genes with a pLI>0.9 and a LOEUF>0.35), and for deleterious missense DNV (8 DNVs compared to 3.2 expected , OR= 2.51, p=0.016, list of 229 genes with a mis-z-score > 3.09). Gene-Set Enrichment Analysis uncovered a highly significant enrichment for genes down-regulated in ME-A cells (breast cancer) undergoing apoptosis in response to doxorubicin (FDR=1.5x10^{-11}) and in fibroblasts expressing mutant forms of ERCC3 after UV irradiation (FDR=5.5x10^{-8}).
Conclusions: We detected an excess of *de novo* variants in CKM, supporting potential pathogenetic mechanisms of disease. Globally, the DNV signal was partially driven by novel constrained genes that are highly expressed during early kidney development. Further gene-set analysis and replication in larger datasets may help pinpoint which genes are most likely driving this enrichment.

PrgmNr 3460 - Complex congenital heart defect in a patient with a pathogenic mutation of *TLK2* gene: Coincidental findings or part of the phenotype of TLK2 syndrome?

View session detail

Author Block: **M. Hajianpour**[1], S. P. Yoon[1], J. L. Carver[1], M. Davis[2]; [1]East Tennessee State Univ. - Quillen Coll. of Med., Johnson City, TN, [2]Ballad Hlth.Med. Associates Family Med. Gray, Johnson City, TN

Disclosure Block: **M. Hajianpour:** None.

Proband is a 23 yo male with TLK2-related condition presenting with learning disability/ mild intellectual disability, and behavioral problems (ADHD, anxiety, socializing problems), congenital heart defects (tetralogy of Fallot, pulmonary artery coarctation and atresia), inguinal hernia, cryptorchidism, and dysmorphic features (dolichocephaly, wide forehead, long face, flat facial profile, malar hypoplasia, small ears, prominent nasal bridge, broad nasal tip, and big/tall chin. **Genetic Work-up:** â¢ Chromosome microarray: An 11.5 Mb region of homozygosity (ROH) at chromosome 22 (26415475_37951451). No specific imprinting disorder known with 22q UPD. â¢ Fragile X study: Negative. â¢ Molecular studies for Invitae Congenital Heart Defects and Heterotaxy Panel (89 genes): A variant of uncertain significance: c.1333C>T (p.Pro445Ser) in SPH4A gene (Its pathogenic mutations are associated with autosomal recessive primary ciliary dyskinesia (PCD): **Non-contributory to phenotype**. â¢ Molecular studies for GeneDx Autism/ID panel (2,500 genes): Heterozygous for a pathogenic variant of TLK2 gene: p.Arg123Ter (CGA>TGA): c.367 C>T in exon 7: **Cause of phenotype** (mother is negative for mutation/ fatherâs test pending). Also hemizygous for a maternally inherited VUS in SMC1A gene: c.1545+5G>A: IVS9+5G>A: **Non-contributory to phenotype.** The TLK2 gene (NM_006852.3) encodes a nuclear serine/threonine kinase involved in DNA checkpoint regulation and chromatin assembly. TLK2 Pathogenic variants cause a neurodevelopmental disorder with mild developmental delay, behavioral concerns, GI problems, and dysmorphic facial features. Additional findings include neonatal feeding difficulties, eye problems, recurrent otitis media, epilepsy, hypotonia, brain malformations, musculoskeletal abnormalities, abnormal skull shape, hypertrichosis, and hoarse voice. Most variants are de novo, but inheritance from affected parents has been reported. A broad spectrum of variants, and a few recurrent variants have been reported. A homozygous missense variant has also been reported in one patient with a more severe phenotype.

The developmental delay, LD/ID, behavioral problems, and dysmorphic features in our patient are compatible with those reported in TLK2 syndrome. However, to our knowledge CHD including TOF and pulmonary coarctation and atresia, as well as inguinal hernia and cryptorchidism have not been reported. The CHD in this patients could be part of the spectrum of TLK2 syndrome (mutation-specific), or be coincidental due to mutation of a gene not included in the Invitae CHD/Heterotaxy panel. No OMIM gene in his 22q ROH is associated with CHD either.

PrgmNr 3461 - Partial trisomy of 8q and partial monosomy of 10q as a result of unbalanced translocation in an infant with multiple congenital anomalies - A Case Report

View session detail

Author Block: **M. Rahman**[1], S. Mikkilineni[2], F. Shakil[1], C. KUMAR[2]; [1]WMC, Valhalla, NY, [2]Westchester Med. Ctr., Valhalla, NY
Disclosure Block: **M. Rahman:** None.
Derivative chromosomes are often caused by unbalanced chromosome rearrangements that appear spontaneously or may be inherited from healthy parents carrying balanced reciprocal translocations. Partial trisomy 8q has been reported sporadically in the literature with variable phenotypes. It has been reported that the isolated duplication of 8q22.2-q24.3 is responsible for dysmorphic features such as hypertelorism, microretrognathia, telecanthus, cleft palate, intellectual disabilities and other phenotypic features. Partial monosomy of 10q is a rare abnormality associated with microcephaly, characteristic facial features, intellectual disability. Here we present a new case of a male infant born at 35 weeks gestation to a 22 year old mother. The infant birth weight was 1.8 kg (~5th percentile), length of 43 cm (~1st percentile) with multiple congenital anomalies, tracheo-bronchomalacia, large ventricular septal defect, tachyarrhythmia, patent ductus arteriosus and respiratory failure. He also has vertebral anomaly, cleft palate, omphalitis and undescended testes. The chromosomal microarray and the cytogenetic investigations were indicated. Chromosome analysis performed on GTG banded metaphases prepared from cultured lymphocytes. The cytogenetic analysis revealed a derivative chromosome 10 with additional unidentified material at 10q region and the infantâ□□s karyotype as 46,XY,add(10)(q?26.1). The microarray analysis performed on whole blood showed a 52 Mb terminal duplication from 8q22.1 to 8q24.3 and a 7.3 Mb terminal deletion from 10q26.2 to 10q26.3. The Fluorescence in situ hybridization (FISH) analysis using 8q and 10q subtelomere probes determined that these abnormalities were associated with an unbalanced translocation between 8q and 10q. The microarray findings revealed that the additional unidentified material at 10q was from the long arm of chromosome 8 representing partial trisomy of that region. In addition, terminal deletion of 10q represents partial monosomy of that region. To the best of our knowledge, this unbalanced translocation between chromosome 8q and 10q is the first report of this combination of chromosomal abnormalities. Finally, this case demonstrates the effective use of cytogenetics, chromosomal microarray and FISH studies for the diagnosis and characterization of an unidentifiable 10q region with complex phenotype for clinical management. It is expected that at least one of the parents is a balanced carrier of t(8;10). Chromosomal analysis of both parents was highly recommended to find out the origin of this abnormality for proper genetic counseling, such as recurrence risk and reproductive options.

PrgmNr 3462 - A rapid and low cost NGS-based cytogenomics platform using proximity ligation technology

View session detail

Author Block: **J. Tapper**[1], S. Eacker[1], I. Liachko[1], S. Sullivan[1], B. Nelson[1], A. Muratov[1], E. Wassman[2], E. Reister[1], M. Van Dyke[1]; [1]Phase Genomics, Seattle, WA, [2]Park City, UT
Disclosure Block: **J. Tapper:** None.
Current cytogenomic strategies for determining the cause of genetic disorders requires an assortment of disparate tests applied in a high-cost, time- and resource-consuming, sequential manner. Each test in the cascade - typically made up of karyotype analysis, chromosomal microarray (CMA), and fluorescence *in situ* hybridization (FISH) - has unique limitations. Karyotype analysis offers a genome-wide view of chromosomal anomalies but has limited resolution. CMA offers increased resolution but is unable to identify balanced translocations, inversions, balanced insertions, and to elucidate complex rearrangements. FISH allows only one (or in some cases a few) loci to be interrogated at a time and requires prior knowledge of the suspected rearrangement. Here we present proximity ligation as an alternative to cytogenetics. This technology offers an approach to identify chromosomal abnormalities that overcomes the afore mentioned challenges of the traditional cytogenetic testing paradigm. Proximity ligation methods have become an industry standard in genome assembly used for the scaffolding and phasing of full-length chromosome sequences. Proof of concept data based on over 50 previously characterized clinical samples shows combining proximity ligation chemistry, off-the-shelf, open-source analytics software and our novel convolutional neural net (CNN) detector achieved sensitivity and specificity comparable to existing cytogenomic tests offered clinically. By leveraging the strengths of proximity ligation sequencing our platform enables the rapid, simultaneous detection of the breadth of chromosomal aberrations obtainable through karyotype analysis, CMA, and every possible FISH probe set in a single, low-cost, highly scalable assay.

PrgmNr 3463 - A Retrospective Review of Dominant Variants Identified from the Horizon™ Carrier Screening Test

View session detail

Author Block: **L. Li**, T. Smart, A. Hoang, D. Li, S. Arun, B. Meyers, J. Gray, B. Gall, N. Sanapareddy, D. Keen-Kim; Natera, Inc., San Carlos, CA
Disclosure Block: **L. Li:** Major Stockholder/Ownership Interest; Natera, Inc. Salary/Employment; Natera, Inc.

The Horizon™ carrier screening test is to help couples assess their risk of passing genetic conditions to their children. Carriers of disease-causing variants are typically healthy and most are without familial history. While carrier screening mostly focuses on variants of recessive inheritance (Antonarakis SE, 2019), dominant variants are also occasionally observed. Dominant variants confer a higher risk of passing genetic conditions to offspring (van Dijk PJ, 2016), and importantly can indicate subclinical conditions or risk of later-onset diseases. Undetected dominant variants may be more common than known in the non-clinical population and their impact on phenotypes remains to be determined (Grzymski JJ, 2020). We aimed to catalogue the P (Pathogenic) or LP (Likely Pathogenic) variants of dominant inheritance within historical Horizon carrier screening data and to evaluate the associated incidental benefit in personal healthcare.

All Horizon carrier screening variants submitted in QCI (QIAGEN Clinical Insight, the clinical decision support software for variant interpretation and reporting) as of April 28, 2021 were retrieved. Variants meeting 2 conditions: 1) Mode of Inheritance assigned as Dominant by QCI , and 2) classified as P or LP by Natera curators were included in the analysis. Evidence of dominant inheritance from publications was checked by manual review.

A total of 185,328 variants were used for this analysis, of which 1421 unique ones have a dominant inheritance pattern marked by QCI. The dominant variants classified as P or LP (N=116) by Natera curators were included in the manual check. The analysis showed that dominant P/LP variants were detected in 11 genes associated with 9 phenotypes: Familial Hypercholesterolemia (LDLR); Alport Syndrome (COL4A3, COL4A4 and COL4A5); Familial Hyperinsulinism (ABCC8); Hypophosphatasia (ALPL); Choroideremia (CHM); Charcot-Marie-Tooth Disease with Deafness (GJB1), Alpha-Thalassemia (HBA1), Beta-Hemoglobinopathies (HBB), and Ornithine Transcarbamylase Deficiency (OTC), with the first two phenotypes being the most commonly observed.

Dominant variants, designated by QCI and confirmed by expert curation, were observed in HorizonTM carrier screening. Long term clinical follow up is important to decipher their penetrance and clinical expressivity. Under the guidance of medical professionals, disease-causing dominant variants could provide an incidental benefit to the carrier screening participants' healthcare, provided that there is no phenotype uncertainty and potential treatment or prevention is available (Miller DT, 2021).

PrgmNr 3464 - Contiguous domains of differentially-accessible, single copy sequences along metaphase chromosomes are conserved among multiple tissues

View session detail

Author Block: **S. Hill**[1], P. K. Rogan[2,3], Y. Wang[1], J. H. Knoll[4,3]; [1]Dept Pathology and Lab. Med., Schulich Sch. of Med. and Dentistry, London, ON, Canada, [2]Schulich Sch Med & Dentistry, London, ON, Canada, [3]Cytognomix Inc., London, ON, Canada, [4]Schulich Sch. of Med. and Dentistry, London, ON, Canada
Disclosure Block: **S. Hill:** None.
Background: During mitosis, chromatin engages in a dynamic cycle of condensation and decondensation. Condensation into distinct units to ensure high fidelity segregation is followed by rapid and reproducible decondensation to produce functional daughter cells. Factors contributing to the reproducibility of chromatin structure between cell generations are not well understood. We investigated local metaphase chromosome condensation along mitotic chromosomes within genomic intervals showing differential accessibility (DA) between homologs (MolCytogen 2014,7[70]; MolCytogen 2015,8[65]). DA was originally identified using sequence-defined single copy (sc) DNA probes with fluorescent in situ hybridization (scFISH) in peripheral lymphocytes (GenomeRes 2001,11:1086; AJMG(A) 2003,121:245). These structural differences between metaphase homologs are non-random, stable, and heritable epigenetic marks which has led to the proposed function of DA as a marker of chromatin memory. Here, we characterize the organization of DA intervals into chromosomal domains by identifying multiple DA loci in close proximity to each other, and examine the conservation of DA between tissues. **Results**: We evaluated multiple adjacent scFISH probes at 6 different DA loci from chromosomal regions 2p23, 3p24, 12p12, 15q22, 15q24 and 20q13 within peripheral blood T-lymphocytes. DA was organized within domains, ranging from 16.0 kb to 129.6 kb in length, based on hybridizations of 2 to 4 scFISH probes per domain extending DA beyond the defined boundaries of individual scFISH probes. Transcriptionally inert chromosomal DA regions in T-lymphocytes also demonstrated conservation of DA in bone marrow and fibroblast cells. **Conclusions**: We identified novel chromosomal regions with allelic differences in metaphase chromosome accessibility and demonstrated that these accessibility differences appear to be aggregated into contiguous domains extending beyond individual scFISH probes and to coincide with previously established topologically-associated domain (TAD) boundaries. DA appears to be a conserved feature of human metaphase chromosomes across different stages of lymphocyte differentiation and germ cell origin, consistent with a role in maintenance of intergenerational cellular chromosome memory.

PrgmNr 3465 - Detection of near constitutional genome-wide uniparental disomy in a patient presenting with Beckwith-Wiedemann syndrome and cholestasis

View session detail

Author Block: **G. J. Fischer**[1,2], M. D. Maxwell[3,4], D. Allingham-Hawkins[1], V. Horner[2], N. Mittal[3], A. Quinn[3]; [1]PreventionGenetics, Marshfield, WI, [2]Wisconsin State Lab. of Hygiene, Madison, WI, [3]Univ. of Texas Hlth.at San Antonio, San Antonio, TX, [4]Univ. Hlth., San Antonio, TX
Disclosure Block: **G.J. Fischer:** Salary/Employment; PreventionGenetics.
Genome-wide uniparental disomy (GW-UPD) is characterized by the inheritance of all 23 sets of chromosomes from a single parent. Constitutional paternal or maternal GW-UPD is lethal *in utero* and, in the case of paternal UPD, presents as hydatidiform mole. However, a small number of individuals with apparent GW-UPD have been reported. To date, all reported individuals are female (as an androgenetic lineage is not viable) and mosaic. In the mosaic state, paternal GW-UPD is compatible with life and affected individuals present with Beckwith-Wiedemann syndrome-like phenotypes and multiple tumors. We describe a female patient born at 31 6/7 weeks gestation who was transferred to the NICU with prematurity, respiratory distress, critical hypotension, and hypoglycemia. Physical exam was consistent with Beckwith-Wiedemann syndrome (BWS). Methylation-specific multiplex ligation-dependent amplification (MLPA) (MRC-Holland, Amsterdam) identified hypermethylation at imprinting center 1 (IC1) and hypomethylation at IC2, confirming a BWS diagnosis. At three months old, the patient developed acholic stools, raising concern for cholestasis. A cholestasis next-generation sequencing panel consisting of 76 genes on DNA derived from whole blood returned only apparently homozygous variant calls, with a read fraction percentage ranging from 90-100%. Since the panel used a whole exome probeset, exome-wide copy number variants and loss of heterozygosity could be evaluated. Loss of heterozygosity was observed across every chromosome, suggesting the possibility of GW-UPD. Confirmatory Illumina SNP microarray (San Diego, CA) testing confirmed high-level mosaic (90% or greater) GW-UPD. Given the previous diagnosis of paternal UPD(11), mosaic paternal GW-UPD was assumed, but not confirmed. At least one report of 93% paternal GW-UPD is documented in the literature of DNA derived from blood, with lower mosaic percentages observed in other tissues (PMID: 29636544). Most reports cite 90% or greater GW-UPD in tumor tissue (PMID: 30882989). To our knowledge cholestasis has not previously been reported in association with paternal GW-UPD. Potential genetic mechanisms for this outcome and changes in clinical management for this patient given this diagnosis will be discussed.

PrgmNr 3466 - Inverted genomic triplication structures: two breakpoint junctions, several possibilities

View session detail

Author Block: **C. M. Grochowski**[1], M. Gandhi[2], H. Du[3], M. G. Mehaffery[2], P. KyungHee[2], J. Eisfeldt[4], M. Pettersson[5], B. Suter[1], S. N. Jhangiani[1], D. M. Muzny[1], J. M. Fatih[3], R. A. Gibbs[1], A. Hastie[6], M. Pendleton[7], E. Harrington[8], S. Juul[7], A. Lindstrand[5], H. Y. Zoghbi[9], J. R. Lupski[3], F. J. Sedlazeck[1], D. Pehlivan[3], C. M. Carvalho[2]; [1]Baylor Coll. Med., Houston, TX, [2]Pacific Northwest Res. Inst., Seattle, WA, [3]Baylor Coll. of Med., Houston, TX, [4]Karolinska Inst, Stockholm, Stockholm, Sweden, [5]Karolinska Inst.t, Stockholm, Sweden, [6]BioNano Genomics, San Diego, CA, [7]Oxford Nanopore Technologies, New York, NY, [8]Oxford Nanopore, New York, NY, [9]Baylor Coll. of Med./HHMI/NRI, Houston, TX
Disclosure Block: **C.M. Grochowski:** None.

Genomic inversions are a class of structural variation (SV) that can occur in a copy-number neutral state or as part of a complex SV with copy number variation (CNV). The duplication-triplication/inverted-duplication (DUP-TRP/INV-DUP) structure is a complex SV formed by replication mechanisms and mediated by a given pair of inverted low-copy repeats leading to two breakpoint junctions *in cis.* The origin of the structure occurs either mitotically as a post-zygotic event or in the male germline. This aberration causes a disease state due to CNVs affecting dosage sensitive genes and is found at a high frequency in several cohorts including *MECP2* duplication syndrome (16%), Pelizaeus-Merzbacher disease (20%), Potocki-Lupski Syndrome and others. Surprisingly, the International Cancer Genome Consortium has recently identified this structure as one of the 12 most prevalent SV mutational signatures found in cancer. We sought to resolve the structure of DUP-TRP/INV-DUP genomic rearrangements in patients, using a combination of array comparative genomic hybridization (aCGH), short-read whole-genome sequencing (WGS), long-read WGS (CRISPR-*Cas9* ONT) and optical genome mapping (OGM). This multiomics approach enabled elucidation of the events within each haplotype, with interpretation of each genomic fragment in context of the larger structure. Astonishingly, from studies of 7 individuals carrying a DUP-TRP/INV-DUP, four distinct and previously unknown subtype structures were identified and could be traced back to the inverted repeat used to mediate the event. In the *MECP2* duplication syndrome samples, the inverted segment varied in size from ~40 kb to ~500 kb and included an amplified copy of *MECP2* in 3/7 patients. Importantly, we can now identify precisely which genomic fragments are inverted within this structure. Although we have previously shown a triplication of *MECP2* presents a more severe phenotype, we are now investigating whether the inverted orientation of *MECP2* leads to transcription-level differences that could contribute to clinical variability. In summary, optical genome mapping and long-read WGS approaches facilitated phasing and assembly of a clinically relevant recurrent SV. These methods have relevance in the study of both Mendelian disease and cancer mutagenesis and confirms our hypothesis that inverted low-copy repeats as a recombinant substrate play a critical role in the underlying mechanism of formation for DUP-TRP/INV-DUP events.

PrgmNr 3467 - Library preparation forâ¯DNAseq: The next generation of automation

View session detail

Author Block: **K. Lu**; Beckman, Indianapolis, IN
Disclosure Block: **K. Lu:** Salary/Employment; Beckman Coulter.
Introductionâ¯ â¯ The ability to sequence genomes rapidly and efficiently has had a significant influence on the scientific community, with applications ranging from basic science toâ¯drug development and translational applications. Creating libraries for next-generation sequencing (NGS) is, unfortunately, a time-consuming operation. Keeping meticulous records of adaptors, as well as consistent pipetting,â¯haveâ¯become crucial for successful library preparation.â¯Many of the processes require precise timing and do not have safe stopping points, leading to a veryâ¯longâ¯workdayâ¯for the user.â¯â¯Methodsâ¯ To overcome these challenges,â¯here weâ¯haveâ¯automatedâ¯the Roche KAPAâ¯HyperPrepâ¯Library Preparation kit onâ¯aâ¯Beckman Coulterâ¯Biomekâ¯NGeniuSâ¯next generation library preparationâ¯system.â¯We loaded the system withâ¯8â¯samples of fragmented DNA, at the input concentrationâ¯ofâ¯100â¯ng.â¯We sequenced theâ¯preparedâ¯librariesâ¯onâ¯anâ¯Illuminaâ¯MiSeqâ¯sequencer and analyzed the data usingâ¯BaseSpace.â¯â¯â¯ â¯Resultsâ¯We created libraries using Roche KAPAâ¯HyperPrepâ¯workflowâ¯and all libraries passed the quality thresholds.â¯Our sequencing results showed thatâ¯>98% of reads were paired. Less than 0.30% of reads were duplicates.â¯ Conclusion:â¯ Automation can create successful libraries from Roche KAPAâ¯HyperPrepâ¯Library Preparation kit.

PrgmNr 3468 - A novel intragenic deletion removing the secondary coiled-coil (CC2) domain of *SOX6* in a patient with neurodevelopmental disorder

View session detail

Author Block: **X. Shao**[1], E. Kellom[2], G. M. Rice[3], V. Horner[4]; [1]Wisconsin State Lab. of Hygiene, Univ Wisconsin Madison, Madison, WI, [2]Waisman Ctr., Univ. of Wisconsin Sch. of Med. and Publ. Hlth., Madison, WI, [3]Univ Wisconsin, Middletown, WI, [4]Univ Wisconsin Madison, Madison, WI
Disclosure Block: **X. Shao:** None.
Background: *SOX6* is a member of the *SOX* (Sry-type HMG box) gene family, which encode transcription factors with a high mobility-group (HMG) domain playing critical functions in many developmental processes, including neurogenesis, sex determination and skeleton formation. Variants in half of the 20 *SOX* genes, including *SOX6*, have been associated with human developmental disorders, called SOXopathies. Known functional domains of SOX6 protein include a C-terminal DNA-binding HMG domain that binds to HMG-like sites in the enhancer region of its regulating genes, and two coiled-coil (CC) homodimerization domains, the primary coiled-coil (CC1) domain and the secondary coiled-coil (CC2) domain, of which the CC1 domain is longer and thought to be more critical than the CC2 domain. SOX6 homodimerizes through the coiled-coil domain and further forms a highly stable complex with DNA.
Case Presentation and Genetic Testing: A 27-year-old male with learning difficulty, developmental delay and short stature was referred to our lab for genetic testing. Microarray analysis followed up with long-range PCR and sequencing analysis identified an 8.2-Kb deletion in *SOX6*. This predicted in-frame deletion removes exons 11 and 12 of *SOX6*, with exon 12 encoding the CC2 domain of SOX6. Though this specific deletion has not been previously reported, larger overlapping intragenic deletions and other sequence variants of *SOX6* have been identified in patients with autosomal dominant Tolchin-Le Caignec syndrome, a neurodevelopmental SOXopathy characterized by mild to severe intellectual disability and developmental delay. This novel intragenic deletion in *SOX6* is classified as likely pathogenic using ACMG guidelines.
Discussion: The current documented intragenic deletions in *SOX6* remove either both coiled-coil domains (CC1 and CC2) or only the CC1 domain, while no *SOX6* variant that deletes only the CC2 domain has been reported. Our study reports the first intragenic deletion variant of *SOX6* that removes solely the CC2 domain in a patient with learning and cognitive delay, and our findings suggest that the CC2 domain of SOX6 is also an essential functional domain playing key roles underlying neurodevelopmental disorders.

PrgmNr 3469 - Analysis of single nucleus sequencing in rat amygdala reveals transcriptional and regulatory mechanisms associated with vulnerability to cocaine addiction

View session detail

Author Block: **J. Zhou**[1], F. Telese[1], G. McVicker[2], A. A. Palmer[3], A. Massarat[1], X. Huang[1]; [1]Univ. of California San Diego, La Jolla, CA, [2]Salk Inst. for Biological Studies, La Jolla, CA, [3]UCSD, La Jolla, CA
Disclosure Block: **J. Zhou:** None.
Misuse of addictive substances inflicts a high societal and financial burden and results in tens of thousands of overdose deaths per year. Thus, there is an urgent need to improve understanding and treatment of substance use disorders (SUDs). Long term drug use can lead to addiction, which is characterized by compulsive drug seeking behaviors and a propensity for relapse even after extended periods of abstinence. This is likely due to widespread neurobiological changes caused by chronic drug use which are known to persist even after cessation of drug use. Twin studies have demonstrated that SUDs are highly heritable, indicating that genetic variation mediates differences in drug response and vulnerability to addiction. However, the mechanisms by which genetic factors drive addiction are not well understood. The majority of genetic risk variants for addiction are found in noncoding genomic regions, making it difficult to determine their gene targets and functional effects. Furthermore, most studies of SUDs have been conducted on heterogeneous bulk tissue and cannot detect cell type-specific signals. This has impeded a comprehensive characterization of the molecular features of the cell types involved in the brainâ□□s reward system and the mechanisms by which chronic drug use leads to neuroadaptations in these cellular circuitries. To address this, we analyzed novel single nucleus RNA-seq and single nucleus ATAC-seq datasets generated from the amygdala of outbred rats from a model of extended access to cocaine self-administration. Our analysis revealed distinct brain cell types and subtypes within which we identified cell type specific differentially expressed genes and differentially accessible open chromatin regions (OCRs) associated with vulnerability to cocaine addiction. We also quantified allele specific expression and open chromatin in each cell type, which we use to boost power for mapping cell type-specific eQTLs and caQTLs, and measured enrichment of risk variants for addiction within cell type specific OCRs. These findings help identify the functional effects of noncoding variants. Finally, we developed a deep learning framework to predict the cell type specific effects of genetic variants on chromatin accessibility. This tool can further elucidate the functional effects of noncoding variants and can be used to assist with selecting causal variants from a set of candidates. In summary, we identified novel cell type specific transcriptional and regulatory mechanisms associated with addiction to advance understanding of the molecular basis of SUDs, which can contribute to more effective treatment approaches and improved patient outcomes.

PrgmNr 3470 - Arimoclomolâ□□s effects on Rimmed Vacuole myopathy caused by *HSPB8* gene mutation

View session detail

Author Block: **E. Mohanty**, L. Weiss, R. Yan, A. Shmara, C. Cheng, V. E. Kimonis; UC Irvine, Irvine, CA
Disclosure Block: **E. Mohanty:** None.
Heat shock protein family B (small) member 8 (HSPB8) is a chaperone protein that is induced by harmful stressful events. HSPB8 acts in conjunction with the co-chaperone BAG3, and the HSPB8-BAG3-HSP70 complex enhances the intracellular clearance of all MND-associated misfolded proteins. Mutation in the HSPB8 gene causes Rimmed Vacuolar Myopathy (RVM), which is characterized by distal myopathy, rimmed vacuoles, and early death. Arimoclomol is a potential treatment for HSPB8 myopathy by acting on the heat shock response of cells. Arimoclomol amplifies and prolongs the activated, HSP-producing state of HSF1. This leads to an amplification in the production of cell protective HSPs in physiologically stressed cells, increasing the amount of activity that helps clear protein aggregates caused by the mutation. The dosages of Arimoclomol tested in patient fibroblasts and iPSC-derived myoblasts, and cell survival by the MTT assay demonstrated an optimal dosage of 4 ÂµM. Western blots of fibroblasts and iPSC-derived myoblasts showed a significant increase in the heat shock proteins and reduction of aggregates. The noticeable increase in HSPB8, HSP70, and BAG3 signify amplification of heat shock response and increased protein aggregate clearance. Arimoclomol with a dosage of 120mg/kg/day is administered through drinking water to Hspb8 mice with patient mutation. The mice have shown to tolerate the treatment well, demonstrating a stable body weight and upward trend of improvement in motor tests. The results have shown that there is potential in using Arimoclomol as a treatment for Rimmed Vacuole myopathy and other myopathic related diseases.

PrgmNr 3471 - Characterization of *APOE* Christchurch carriers in 239,390 participants of the UK Biobank with whole exome and/or whole genome sequencing data

View session detail

Author Block: **K. Y. He**[1], E. A. Khramtsova[1], A. Cabrera-Socorro[2], L. De Muynck[2], Q. S. Li[3], B. Smets[2], A. Parrado[1], B. A. J. Sarver[1], H. A. Hejase[1], I. Royaux[4], S. Li[1], M. Black[1], S. Lovestone[2]; [1]Computational Sci., Janssen R&D, Spring House, PA, [2]NeuroSci. Therapeutic Area, Janssen R&D, Beerse, Belgium, [3]NeuroSci. Therapeutic Area, Janssen R&D, Titusville, NJ, [4]Functional Genomics, Computational Sci., Beerse, Belgium

Disclosure Block: **K.Y. He:** Salary/Employment; Janssen R&D.

An ultra-rare *APOE* Christchurch (APOECh) mutation (c.460C>A or p.Arg154Ser) has been hypothesized to protect from *PSEN1*-based autosomal dominant Alzheimerâ□□s disease (AD). The protective effect was observed for a single homozygous individual but not for heterozygous descendants. Presence of the APOECh mutation in heterozygous carriers is associated with dyslipidemia and premature cardiovascular disease (CVD). To date, few studies have reported on the APOECh. The gnomAD v3.1 database includes 7 heterozygous individuals in 76,065 whole genome sequencing (WGS) samples (allele frequency [AF] = $4.6x10^{-5}$), 3 of whom are of Non-Finnish European ancestry and 4 of Latino/Admixed American ancestry. To study the phenotypic spectrum of APOECh carriers, we examined the UK Biobank (UKB) whole exome sequencing (WES; N=200,643) and WGS (N=141,948) data. We observed 14 heterozygous carriers of APOECh in WES (AF=$3.5x10^{-5}$) and 7 heterozygous carriers (AF=$2.5x10^{-5}$) in WGS, with 5 individuals present in both; AF in UKB is consistent with that reported in gnomAD. The 16 unrelated individuals include 12 females and 4 males with last known age ranging from 50 to 72; 15 are European and 1 is Admixed American by genetic ancestry assessment. None of these individuals carry Mendelian mutations in *APP*, *PSEN1*, and *PSEN2* for early-onset AD, and only 1 is a heterozygous carrier of the ApoE4 risk allele. Upon examining ICD-coded clinical history, self-reported illnesses, and parental medical history, none of these individuals have any form of dementia or family history of dementia. There are 6 participants with history of CVD and 5 show abnormal or borderline abnormal lipid biomarker levels according to AHA guidelines, among which 1 is deceased due to acute myocardial infarction, with data not reported in 1 individual. Additionally, there are 4 participants with abnormal lipid biomarkers levels with no reported history of CVD. Eleven individuals have self-reported medication information, 7 have primary care prescription records, and 5 have both sources of information available. No therapies for dyslipidemia were reported for these 16 individuals; the most commonly used drug is proton pump inhibitor (PPI; N=7) and 5 of these users have reported gastrointestinal (GI) disorders. The heterozygous APOE4/APOECh carrier has a medical history of GI neoplasms and PPI use. While the APOECh variant is very rare and larger cohorts are needed to assess its contribution to dementia, dyslipidemia, and CVD, the UKB provides an unprecedented opportunity to follow these carriers and elucidate the underlying role of APOECh in disease etiology.

PrgmNr 3472 - Drosophila functional screening of de novo variants in autism uncovers deleterious variants and facilitates discovery of rare neurodevelopmental diseases

View session detail

Author Block: **P. C. Marcogliese**[1], S. L. Deal[1], J. Andrews[1], J. Harnish[1], V. Bhavana[1], H. K. Graves[1], S. Jangam[1], X. Luo[1], N. Liu[1], Y-H. Chao[1], B. Hull[1], P-T. Lee[1], C. M. Longley[1], H-T. Chao[1], H. Chung[1], N. A. Haelterman[1], L. Z. Rossetti[1], A. Gerard[2], E. Sukarova-Angelovska[3], S. Trajkova[4], B. Keren[5], L. Perrin[6], T. Courtin[7], F. Tran Mau-Them[8], J. DELANNE[9], E. Rosenhahn[10], K. Platzer[11], V. Strehlow[10], E. M. C. Schwaibold[12], R. Guerrini[13], A. Vetro[13], E. England[14], C. N. Murali[1], S. Barakat[15], M. F. van Dooren[16], M. Wilke[17], M. A. van Slegtenhorst[18], G. Lesca[19], I. Sabatier[20], N. Chatron[21], J. A. Madden[22], C. A. Brownstein[23], P. B. Aggrawal[22], R. Keller[24], L. Pavinato[25], A. Brusco[26], J. A. Rosenfeld[27], R. Marom[1], M. F. Wangler[28], S. Yamamoto[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]Univ. Clinic for Children's Diseases, Skopje, Macedonia, The Former Yugoslav Republic of, [4]Univ. of Turin, Torino, Italy, [5]PitiÃ©-SalpÃªtriÃ¨re Hosp., APHP.Sorbonne UniversitÃ©, Paris, France, [6]Robert DebrÃ© Hosp., Paris, France, [7]Sorbonne UniversitÃ©s, UPMC UniversitÃ© Paris, Paris, France, [8]CHU Dijon Bourgogne, Dijon, France, [9]Ctr. Hosp.ier universitaire de Dijon, Dijon, Dijon, France, [10]Univ. of Leipzig Med. Ctr., Leipzig, Germany, [11]Leipzig, Germany, [12]Heidelberg Univ., Heidelberg, Germany, [13]Childrenâ□□s Hosp. Meyer-Univ. of Florence, Florence, Italy, [14]The Broad Inst. of MIT and Harvard, Cambridge, MA, [15]Wissenkerke, Netherlands, [16]Erasmus MC, Rotterdam, Netherlands, [17]Erasmus MC Univ. Med. Ctr., Rotterdam, Netherlands, [18]EUR Med. Ctr., Rotterdam, Netherlands, [19]Lyon Univ. Hosp., UniversitÃ© Claude Bernard, Lyon, France, [20]Lyon Univ. Hosp., Lyon, France, [21]Service de GÃ©nÃ©tique, Hospices Civils de Lyon, Bron Cedex, France, [22]Boston Childrenâ□□s Hosp., Boston, MA, [23]Boston Children's Hosp., Boston, MA, [24]Adult Autism Ctr., Turin, Italy, [25]Univ. of Turin, Turin, Italy, [26]Univ. of Torino, Torino, TO, Italy, [27]Houston, TX, [28]Baylor Coll. Med., Houston, TX

Disclosure Block: **P.C. Marcogliese:** None.

Individuals with autism spectrum disorders (ASD) exhibit an increased burden of de novo variants in a broadening range of genes. We functionally tested the effects of ASD missense variants using Drosophila through â□□humanizationâ□□ rescue and overexpression-based strategies. We studied 79 ASD variants in 74 genes identified in the Simons Simplex Collection and found 38% of them caused functional alterations. Moreover, we identified GLRA2 as the cause of a spectrum of neurodevelopmental phenotypes beyond ASD in eight previously undiagnosed subjects. Functional characterization of variants in ASD candidate genes point to conserved neurobiological mechanisms and facilitates gene discovery for rare neurodevelopmental diseases.

PrgmNr 3473 - Evaluation of chromosomal copy numbers in the 5â□□UTR of *MBD5*

View session detail

Author Block: **S. V. Mullegama**[1], S. Milosavljevic[2], F. J. Probst[3], J. W. Innis[4], R. Mendoza-Londono[5], T. Ezashi[6], S. H. Elsea[7]; [1]Sam Houston State Univ., Willis, TX, [2]Molecular & Human Genetics, Baylor Coll. Med., Houston, TX, [3]Houston, TX, [4]Univ Michigan, Ann Arbor, MI, [5]Hosp for Sick Children, Toronto, ON, Canada, [6]Div. of Animal Sci. and Bond Life Sci. Ctr., University of Missouri, MO, [7]Baylor Coll. Med., Houston, TX

Disclosure Block: **S.V. Mullegama:** None.

Diagnostic genomic tools, specifically chromosomal microarray analysis and whole exome sequencing (WES) allow for detection of copy number variants (CNVs) or single nucleotide variants (SNVs) in patients with genetic disorders. While these tools are powerful, due to the inherent nature of these tests, they neglect an important region of a gene: the 5â□□untranslated region (5â□□UTR) which is key in the regulation of gene transcription. We hypothesize that it is plausible that CNVs or loss of function SNVs in the 5â□□UTR of dosage-sensitive genes are the underlying cause of the phenotype of the patient due to impaired or altered transcription. To give credence to this hypothesis, we utilized our *MBD5*-associated neurodevelopmental disorder (MAND) patient population. MAND is an umbrella term that that describes a group of disorders that include 2q23.1 deletion syndrome, 2q23.1 duplication syndrome, and *MBD5* SNVs that affect the function of methyl-binding domain 5 (MBD5), and share a common set of cognitive and autism-like behaviors. We recruited a cohort of MAND patients with deletions in the 5â□□UTR region of *MBD5* (**n=44**) to illustrate whether deletions in the 5â□□UTR region of *MBD5* may be the etiology responsible for haploinsufficiency of *MBD5* which is seen in all 2q23.1 deletion patients. Through genotype-phenotype correlation studies, RNA-seq analysis, and mRNA expression analysis, we illustrate that there are specific 5â□□UTR deletions in *MBD5* which result in haploinsufficiency and produce solely an intellectual disability (ID) and autistic phenotype suggesting that deletions in the 5â□□UTR of *MBD5* are a hot spot for ID and autism. Further, our study revealed a possible alternative transcript of MBD5 with a transcriptional start site in the 5â□□UTR of the primary transcript that may impact disease. Thus, it is conceivable that deletions of the start site for the alternative transcript leads to loss of MBD5 function in some cells critical for neurodevelopment. Finally, our study points to the necessity of investigating the 5â□□UTR in dosage-sensitive genes to fully explore the potential for disease-causing variants. In conclusion, the coverage of the 5â□□UTR in diagnostic tests is imperative and requires test development to gain near 100% coverage of the 5â□□UTR especially in dosage-sensitive genes.

PrgmNr 3474 - Functional characterization of 400 acid Î±-glucosidase variants and analysis of phenotype-genotype relationships in Pompe disease

View session detail

Author Block: **R. Mishra**[1], D. Kanavy[2], C. Bakolitsa[1], J. L. Goldstein[2], Z. Hu[1], D. S. Bali[3], Y. Ru[4], G. K. Yu[5], J. H. LeBowitz[6], W. T. Clark[7]; [1]Univ. of California, Berkeley, Berkeley, CA, [2]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Duke Univ., Durham, NC, [4]Atara Biotherapeutics, South San Francisco, CA, [5]Global Blood Therapeutics, South San Francisco, CA, [6]Novato, CA, [7]Biomarin Pharmaceutical, Novato, CA

Disclosure Block: **R. Mishra:** None.

Pompe disease (PD) is an autosomal recessive metabolic disorder caused by pathogenic variants in the lysosomal enzyme acid Î±-glucosidase *(GAA)*, leading to partial or complete loss of *GAA* activity. The resulting, progressive accumulation of glycogen in different tissues presents with a broad range of clinical manifestations that differ by age of onset (Infantile, Juvenile, and Adult), severity and disease progression. Timely diagnosis of PD involves detection of *GAA* deficiency and *GAA* sequence analysis, and is critical so that therapy can be initiated to prevent early mortality. However, the number of rare variants of unknown significance (VUS) identified in *GAA*, and lack of functional data for these variants, complicates their interpretation as well as the prediction of age of onset and severity for babies identified at birth. Here we describe the largest single functional study of *GAA* missense variants to date (n = 400). We apply the weight of evidence for ACMG-AMP functional criteria codes PS3 and BS3, thus facilitating the clinical interpretation of *GAA* variants. We observed that 15/16 (94%) of tested variants annotated as pathogenic or likely-pathogenic in ClinVar had activity levels below 15% wild-type activity. We found that 94/364 (26%) of tested VUS (absent, or annotated as having uncertain significance or conflicting interpretations of pathogenicity in ClinVar) reduce *GAA* activity to levels similar to those observed in known disease-causing mutations and may have possible disease relevance. We also curated PD patient genotype and age of onset data from the literature (n = 818 patients). We apply a combinatorial phenotype matrix previously presented for metachromatic leukodystrophy, that, given a patientâ□□s alleles, predicts disease severity and onset. We made several observations for PD, the most striking of which was that there are several genotype combinations consisting of two pathogenic, although mild in impact, alleles that may not lead to a discernible disease phenotype. Overall, it is hoped that this work will advance the understanding of genotype-phenotype correlations in PD, helping genetic counseling and future therapies in families affected with the disease.

PrgmNr 3475 - Genetic regulation of gene co-expression in human prefrontal cortex

View session detail

Author Block: **A. Cote**, H. Young, L. M. Huckins; Icahn Sch. of Med. at Mount Sinai, New York, NY
Disclosure Block: **A. Cote:** None.
Background: Organizing disease-relevant variants and genes by shared biological relationships provides an informative additional strategy to investigate the systems and pathways underlying genetic risk for psychiatric disorders. Moving from a gene to systems-level interpretation of GWAS findings is an important next step to define the regulatory networks and cell types implicated in psychiatric disorders, and improve reproducibility in the context of allelic and locus heterogeneity. Current eQTL methods are limited in their ability to characterize the effects of variants on distant genes. While cis-eQTLs have been thoroughly characterized in the prefrontal cortex, trans-eQTL discovery has been limited due to the considerable multiple testing burden of testing for relationships between every variant and gene in the genome. However, trans-eQTLs can also provide key insights into how GWAS variants exert their effect. By studying groups of co-expressed genes, we overcome the strict significance threshold of a trans-eQTL search; co-expression modules are leveraged for dimensionality reduction and allow for the study of distant variant effects on gene co-regulation.**Results:** The primary analysis was conducted in 491 individuals of European ancestry from the CommonMind Consortium dataset**.** 715 total co-expression modules were identified through multiscale embedded gene co-expression network analysis (MEGENA). Modules contained an average of 72.2 genes (SD=187.2). Significant associations between single nucleotide polymorphisms (SNP) and the first principal component of a co-expression module (i.e. co-eQTLs) were identified for 107/715 modules at a per module Bonferroni threshold of p Conclusion: Our results show successful identification of SNPs associated with changes in gene co-expression for 15% of tested co-expression modules, representing both cis and trans SNP-gene relationships. This project aims to determine the effect of common genetic variants on gene co-regulation in the prefrontal cortex, in a targeted effort to assign novel functional roles to psychiatric disorder GWAS risk loci and promote the development of improved diagnostics and therapeutic targets.

PrgmNr 3476 - Identifying novel evolutionarily conserved regulators of dopamine metabolism using a forward genetic screen in Drosophila melanogaster

View session detail

Author Block: **S. Deal**[1], E. S. W. Seto[1,2], D. Bei[1], S. Gibson[1], S. Yamamoto[3]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]Baylor Coll. of Med., HOUSTON, TX
Disclosure Block: **S. Deal:** None.
Neuronal communication is indispensable for proper brain function, and neurotransmitters are a core signaling component for this process. Dysregulation of the neuromodulator dopamine (DA), for example, leads to social and behavioral impairments in animal models and neurological and neuropsychiatric disorders in humans. Understanding the cellular and molecular mechanisms that regulate DA synthesis, secretion and metabolism is paramount to finding better treatments for patients with DA-related disorders. Moreover, identifying novel regulators of DA can provide a list of candidate genes for human neurological and psychiatric disorders.
Since core components of DA synthesis and secretion are highly conserved, we performed a forward genetic screen in *Drosophila melanogaster*. Using tissue-specific RNAi and chemical mutagenesis/clonal analysis, we identified 123 genes that affect cuticle pigmentation, a DA-dependent phenotype in insects. 85% of these genes are conserved between flies and humans. Knockdown of many of these genes in DArgic neurons led to alterations in basal activity measured by Drosophila Activity Monitors (DAM), suggesting a role for them in DA regulation in the brain as well as epidermal cells. A cohort of 23 genes from the cuticle screen was selected for further DA measurement using High Performance Liquid Chromatography (HPLC) based on strength of phenotypes and neurological disease association. We observed a significant change in total head dopamine in 5 fly genes corresponding to 8 human genes: *AP-1*

PrgmNr 3477 - Interrogating strategies to link CRISPR screens to neurodegenerative disease

View session detail

Author Block: **B. van de Geijn**[1], E. Peguero Sanchez[2], V. Ramaswamy[1], J-P. Fortin[1], M. McCarthy[1], T. R. Bhangale[1]; [1]Genentech, South San Francisco, CA, [2]Roche, Warsaw, Poland
Disclosure Block: **B. van de Geijn:** Salary/Employment; Genentech.
Genome-wide pooled CRISPR screening has emerged as a powerful tool to elucidate the mechanisms underlying disease genetics. Integrating these screens with genome-wide association studies (GWAS) can validate relevance to diseases of interest, identify causal genes/cell-types, and identify disease-relevant processes. However, many methodological challenges remain. To address these, we identify best strategies to: link screens to disease, integrate protein-protein interaction networks, known pathways, and account for essential genes. We demonstrate the potential of the approach by comparing 52 screens against GWAS for four neurodegenerative diseases: Parkinsonâ□□s disease, Alzheimerâ□□s disease, Multiple Sclerosis, and Amyotrophic Lateral Sclerosis.

We determine the optimal strategy for linking genesets to disease (gs2d) with GWAS. We test strategies based on: stratified LD score regression (Finucane et al. Nat Gen 2015), MAGMA (de Leeuw et al. PLoS Comp Bio 2015), and fine-mapping (Wang et al. RJSS 2020). While GWAS results are at variant level, CRISPR screen results are typically at gene-level. Since linking variant-to-gene (v2g) remains a substantial challenge, we evaluate various options: windows around genes (Finucane et al Nat Gen 2015), activity-by-contact (Fulco et al. Nat Gen 2019, Jagadeesh et al. biorXiv), and expression modification prediction (Wang et al. medRxiv). To establish a standard set of positive controls, we match Online Mendelian Inheritance of Man (OMIM) catalog genesets with corresponding disease GWAS summary statistics. We assess each strategy for power and calibration under ideal, noisy, and control conditions. Based on initial comparisons, we find MAGMA strategies significantly outperform the others.

We apply the best gs2d strategy to screens from the CRISPR brain project (Tian et al. Nat Neuro 2021) as well as neurodegenerative disease related screens we conducted. We reprioritize the original screen results using random walk with restart on protein-protein interaction networks, which has been shown to improve geneset enrichment (Fang et al. Nat Gen 2018). We then identify enriched pathways within these reprioritized sets. Finally, we link genes from each screen and enriched pathway to disease. We find statistically significant associations including those between Parkinsonâ□□s disease (PD) and neuronal-specific lysosomal function, lipid homeostasis, oxidative stress, and ferroptosis. We thus provide human genetic evidence for the role of these processes known to be involved in PD. In summary, we provide a framework for integrating screens with disease while testing the best strategies for each piece.

PrgmNr 3478 - Lessons learned from genomics studies of a large recessive NDD cohort: High prevalence of multilocus pathogenic variation, candidate novel genes, and extreme genetic heterogeneity

View session detail

Author Block: **D. Pehlivan**[1,2], T. Mitani[1], S. Isikay[3], A. Gezdirici[4], E. Yilmaz Gulec[5], J. M. Fatih[1], I. Herman[6], D. Calame[6], G. Akay[1], H. Du[6], C. M. Grochowski[7], A. Jolly[8], M. Dawood[1], R. DUAN[9], S. N. Jhangiani[7], D. Marafi[10], Z. H. Coban Akdemir[11], C. M. Carvalho[12], R. A. Gibbs[7], J. E. Posey[1], J. R. Lupski[13]; [1]Baylor Coll. of Med., Houston, TX, [2]Texas Children's Hosp., Houston, TX, [3]Univ. of Gaziantep, Gaziantep, Turkey, [4]Kanuni Sultan Suleyman Training and Res. Hosp., ISTANBUL, Turkey, [5]Kanuni Sultan Suleyman Training and Res. Hosp., Istanbul, Turkey, [6]Houston, TX, [7]Baylor Coll. Med., Houston, TX, [8]1 Baylor Plaza, Houston, TX, [9]HOUSTON, TX, [10]Kuwait Univ., Safat, Kuwait, [11]UTHlth.Sch. of Publ. Hlth., Houston, TX, [12]720 Broadway, Seattle, WA, [13]Baylor Col Med, Houston, TX

Disclosure Block: **D. Pehlivan:** None.

Neurodevelopmental disorders (NDDs) represent the largest category of Mendelian disease traits with a wide array of features including developmental delay/intellectual disability, structural brain anomalies, epilepsy and neurobehavioral/neuropsychiatric problems. Concomitantly, >80% of all human genes are expressed during brain development and/or are critical for brain function. Despite the surge in gene discovery fostered by next generation sequencing, the majority of NDD cases remain molecularly undiagnosed. Our previous exome sequencing (ES) in 128 Turkish NDD families (TBM1) yielded 41 candidate NDD genes. Of note, 6 out of 19 of genes, which we considered to have phenotypic expansion were found to have deleterious variants in an additional locus to that of the primary molecular diagnostic locus/gene (i.e. multilocus pathogenic variation, MPV). We observed that MPVs, mainly driven by identity-by-decent (IBD) causing â□□genomic blocksâ□□ of Absence of Heterozygosity (AOH), are present in much higher ratios in this cohort, and hypothesized that IBD may represent a genetic/genomic mechanism for intra- and interfamilial variability, especially in admixed populations with an elevated coefficient of consanguinity. To investigate this hypothesis, ES and whole genome sequencing (WGS) were now conducted in a second cohort consisting of 254 Turkish NDD kindreds [234 newly enrolled (TBM2) and 20 families whom remained unsolved from TBM1]. We implemented family-based rare variant analyses using internally established stringent variant prioritization criteria with in-house developed (HMZDelFinder and NMDEscPredictor) and/or publicly available bioinformatic tools (e.g. xAtlas).A molecular diagnosis was rendered in 181 out of 254 families (75%). Of note, out of 181 â□□solvedâ□□ families, deleterious variants were identified in 226 distinct genes. We propose 86 genes in which deleterious variants are candidate causative for an NDD phenotype. Importantly, based on objective variant prioritization criteria, we identified 54 families (54/181=29%) with MPV, mostly driven by AOH due to IBD. We established molecular diagnoses in 5/20 families that remained â□□unsolvedâ□□ in the TBM1 cohort mainly using additional bioinformatic tools and studying expanded family members. Our study shows the astonishing contribution of MPV in Mendelian disease traits and also underscores the tremendous genetic heterogeneity of nervous system development and function. The study also contributes several novel â□□candidate NDD genesâ□□ and further demonstrates the diagnostic utility of reanalysis in unsolved cases.

PrgmNr 3479 - Modeling the Impact of Alzheimerâ□□s Disease Genetic Risk on Microglia States and Functions

View session detail

Author Block: **M. Therrien**[1], M. Dolan[1], S. Jereb[1], T. Kamath[1], S. Marsh[1], M. Johnson[1], E. Macosko[1], K. C. Eggan[2], B. Stevens[1]; [1]Broad Inst., Cambridge, MA, [2]Stowers Med. Inst/HHMI, Cambridge, MA
Disclosure Block: **M. Therrien:** None.
Microglia, the brainâ□□s resident immune cells, are highly dynamic and reactive to environment and genetic challenges. More than 40 genomic loci have been linked to late onset Alzheimerâ□□s disease (AD) and many risk genes are highly expressed in microglia. Our goal is to connect insights from genetic association studies to new ways of functionally modeling the cellular and molecular causes of disease to enable predictive tracking and targeting of detrimental immune cell states in patients in the early stages of disease. Single cell transcriptomic studies reveal diverse microglial states in human and mouse brains, however we currently lack specific approaches to track and manipulate specific populations of microglia in Alzheimerâ□□s and other diseases. To answer this question, we turned toward human iPSC-induced microglia (iMGL) and single cell transcriptomics.Single cell transcriptomics revealed the presence of multiple microglia states (up to 17) when iMGLs are stimulated with different brain-relevant challenges, including apoptotic neurons, synaptosomes, myelin and amyloid. Moreover, alignment of these data using Liger (Welch et al. Cell 2017) shows these states are similar to the ones observed in human and mouse in vivo, revealing several disease associated states, including disease -associated microglia (DAM). We also observed changes in microglia states depending on the challenge and genetic background.Together, our data identified key elements causing the formation of DAM and how AD risk genes affect disease -associated states and functions. This work will open the door to the identification of modulators of DAM and highlight new therapeutics avenues of AD.

PrgmNr 3480 - Molecular Dynamics Simulations and Markov State Models Reveal Metastable States and Isoform-Dependent Kinetics in Apolipoprotein E Risk and Resilience Alleles

View session detail

Author Block: **W. Martin**, F. Cheng; Cleveland Clinic, Cleveland, OH
Disclosure Block: **W. Martin:** None.

Apolipoprotein E (ApoE) is the primary cholesterol and lipid transporting apolipoprotein in the central nervous system (CNS). Human genetics studies have demonstrated that the Îµ4 isoform, which differs from the common Îµ3 isoform only by a substitution of an arginine for cysteine at position 112, is the highest genetic risk factor for late-onset Alzheimer Disease (AD), while the Îµ2 isoform, differing only by a cysteine at position 158 as opposed to arginine, is protective. Further, the ApoE Îµ3 Christchurch mutation has been proposed to confer resistance to autosomal dominant AD. Evidence supports a direct interaction with AÎ² oligomers as well as colocalization with oligomers near AÎ² plaques. While significant progress has been made in investigating the properties of ApoE, there is more to be uncovered regarding how these single amino acid variants so drastically impact AD risk. It has previously been proposed that a salt bridge between Arg 61 and Glu 255 is key to the structural difference in Îµ4 due to the arginine at position 112 forcing Arg 61 into a more solvent-accessible direction, while a cysteine at 112 allows Arg 61 to remain within the helix bundle. Mutation of Arg 61 to threonine (or Glu 255 to alanine) â□□rescuedâ□□ Îµ4, causing it to function similarly to Îµ3. However, the solved structure of a monomeric mutant of Îµ3, which has been demonstrated to be functionally equivalent to the wild type, did not have such an interaction present; instead, an Glu 255 interacts with Lys 95 and is spatially distant from Arg 61. Additionally, it has been proposed that a significant conformational transformation is required upon lipidation, with most models proposing an extended conformation whereby the protein wraps around the hydrophobic tails of a lipid disc. To date, the monomeric mutant is the only full-length solved structure for ApoE. This leaves unanswered questions regarding the supposed conformational differences between the isoforms. Here, we have implemented long-timescale molecular dynamics simulations (totaling 100 microseconds, 5 replicates at 2 Î¼s each and a single 15 Î¼s simulation for each isoform) and Markov state models to investigate the isoform-dependent differences in the molecular kinetics of ApoE. The isoform-specific metastable states elucidated by our simulations could shed light on why each isoform interacts differently with AÎ², has a preference for different sized lipid discs, and the impact of Arg 61 on the intramolecular interactions in the Îµ4 isoform. This study could reveal conformational states which may be modulated by small-molecule structure correctors, potentially not only ameliorating the risk conferred by Îµ4 with respect to Îµ3, but possibly Îµ2.

PrgmNr 3481 - Mutations Near the N- or C-Terminus of AHDC1 are Associated with More Mild Xia-Gibbs Syndrome Phenotypes

View session detail

Author Block: **J. Hu**[1], V. Chander[1], T. Balci[2], K. Van Engelen[2], H. J. Vernon[3], R. McClellan[4], M. M. Khayat[1], M. Dawood[1,5,6], J. E. Posey[5], J. R. Lupski[1,5,7,8], D. R. Murdock[1,5], M. F. Wangler[5], R. A. Richard[1,5]; [1]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [2]Div. of Genetics, Dept. of Paediatrics, Western Univ., London, ON, Canada, [3]Dept. of Genetic Med., Johns Hopkins Univ., Baltimore, MD, [4]Kennedy Krieger Inst., Baltimore, MD, [5]Dept. of Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX, [6]Med. Scientist Training Program, Baylor Coll. of Med., Houston, TX, [7]Texas Children's Hosp., Houston, TX, [8]Dept. of Pediatrics, Baylor Coll. of Med., Houston, TX
Disclosure Block: **J. Hu:** None.
Xia-Gibbs Syndrome (XGS; MIM# 615829) was initially described in four severely affected individuals with *de novo* mutations in the 1603 amino acid AT-Hook DNA Binding Motif Containing 1 gene (*AHDC1*), leading to predicted truncated protein synthesis. In just four years, our HIPAA-compliant XGS Registry has identified 280 sequence-confirmed XGS families worldwide and accrued detailed clinical records and research consents from 90. Through the XGS Registry we have: i) documented XGS clinical heterogeneity and that most XGS individuals are severely affected, requiring life-long care; ii) identified core XGS phenotypes (i.e. a clinical synopsis) including speech delay, intellectual disability (ID), developmental delay and hypotonia (PMID 29696776); iii) associated the risk of secondary XGS features of scoliosis and seizure with the position of the *AHDC1* mutation (PMID 33644933); iv) identified XGS patients with missense mutations in *AHDC1,* where mutation position relative to functional domains was a key factor supporting their diagnosis (*Submitted for publication*) and; v) resolved putative XGS diagnoses in individuals with large genomic deletions (*in preparation*). Here, we describe ten XGS individuals harboring truncating variants with mild phenotypes, challenging the established paradigm that all *AHDC1* truncation mutations result in â□□severeâ□□ XGS. A 14 year old male being evaluated for epilepsy was identified with a paternally inherited *AHDC1* truncating mutation at codon 108. This XGS father and child have mild learning disabilities, modest memory loss and sub-clinical sleep apnea. An unrelated 38 year old individual, with a *de novo* stop-gain mutation at codon 136, obtained a college degree and employment. Two 15 year old identical twin females with a *de novo AHDC1* frameshift beginning at codon 1325 have very mild dysmorphism, mild/moderate autism spectrum disorder and hypotonia, but function with good conversational language skills. The remaining XGS putatively-mild individuals are less than 10 years old, complicating the assessment of their long-term clinical outcomes, but many have the ability to speak more than 200 words and a diagnosis of autism, rather than ID.
Together these observations support a â□□polarity of mutation effectâ□□ model wherein truncating mutations close to either the N- or C-terminus of the AHDC1 protein may lead to an overall mild XGS clinical outcome. Further, the data indicate the power of the patient Registry to provide insights into genotype-phenotype correlations, and the value of expanding both the number of registrants and depth of clinical feature ascertainment over the natural history of individuals with XGS.

PrgmNr 3482 - Personalized structural biology reveals the molecular mechanisms underlying heterogeneous epileptic phenotypes caused by *de novo* KCNC2 variants

View session detail

Author Block: **S. Mukherjee**[1], T. Cassini[2], B. Li[1], N. Hu[3], T. Yang[3], W. Shen[3], C. W. Moth[1], D. C. Rinker[1], J. H. Sheehan[4], Undiagnosed Diseases Network, R. L. Macdonald[3], D. M. Roden[3], J. Meiler[5,1,6], G. Kuenze[6], J. A. Phillips[3], J. A. Capra[7,1]; [1]Vanderbilt Univ., Nashville, TN, [2]Chevy Chase, MD, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Washington Univ. Sch. of Med., St. Louis, MO, [5]Leipzig Univ., Leipzig, Germany, [6]Leipzig Univ. Med. Sch., Leipzig, Germany, [7]Univ. of California San Francisco, San Francisco, CA

Disclosure Block: **S. Mukherjee:** None.

Background. DNA sequencing is now common in diverse clinical contexts, from rare diseases to cancer. However, sequencing often fails to provide actionable insight, due to the challenge of proposing and validating testable hypotheses about the effects of variants of unknown significance (VUS). Here, we describe a â□□personalized structural biologyâ□□ (PSB) approach that addresses this challenge by leveraging recent innovations in the determination and analysis of protein 3D structures.
Results. We illustrate the power of the PSB approach in an individual from the Undiagnosed Diseases Network (UDN) with developmental epileptic encephalopathy (DEE) symptoms who has a novel *de novo* VUS in *KCNC2* (p.V469L), the Kv3.2 voltage-gated potassium channel. A nearby *KCNC2* variant (p.V471L) was recently suggested (without functional evidence) to cause DEE-like phenotypes. We find that both variants are located in the conserved hinge region of the S6 helix and likely to affect protein function; however, despite their proximity, computational structural modeling suggests that the V469L variant is likely to sterically block the channel pore, while the V471L variant is likely to stabilize the open state. Biochemical and electrophysiological analyses demonstrated heterogeneous loss-of-function and gain-of-function effects, respectively. Using computational structural modeling and molecular dynamics simulations, we revealed the molecular basis for these distinct effects.
Conclusions. Our results implicate *KCNC2* as a causative gene for DEE and delineate the molecular basis for the heterogeneous clinical phenotypes for the two proximal variants. This demonstrates how the PSB approach provides a broadly applicable analytical framework for individualized hypothesis-driven interpretation of protein-coding VUS suspected to contribute to disease.

PrgmNr 3483 - *MPDZ*, a potential causal gene for the bone mineral density association at chromosome locus 9p23, regulates *in vitro* osteoblast function and *in vivo* bone mass accrual

View session detail

Author Block: **A. Buo**[1], J. Stains[2], C. R. Farber[3]; [1]Univ. of Virginia, Charottesville, VA, [2]Univ. of Maryland, Baltimore, Baltimore, MD, [3]Univ Virginia, Charlottesville, VA
Disclosure Block: **A. Buo:** None.
Osteoporosis is a highly debilitating disease of low bone mass and increased fracture risk. To combat the rising prevalence of the disease, research endeavors have increasingly focused on elucidating the underlying genetic determinants modulating bone health. Genome-wide association studies (GWASs) have proven to be very effective in this regard, collectively discovering hundreds of genetic loci that are associated with changes in bone mineral density (BMD). However, the identities of the variants and genes that are causal for many of these BMD GWAS associations have yet to be determined. One locus, located at human chromosome 9p23, contains a lead variant (rs12340775; $P = 3.8 \times 10^{-10}$) that resides in intron 3 of the Multiple PDZ Domain Crumbs Cell Polarity Complex Component (*MPDZ*) gene. Utilizing data from the Genotype-Tissue Expression (GTEx) consortium, we observe that this BMD variant is also a colocalizing expression quantitative trait locus (eQTL) for *MPDZ* expression in cultured fibroblasts, thereby establishing a possible causal link between differences in *MPDZ* expression and the BMD association at this locus. Here, we present initial evidence for a novel role for *MPDZ* in the regulation of bone mass accrual in mice, thus supporting its involvement in driving the genetic BMD association at 9p23. Given that *MPDZ* is expressed in osteoblasts and not osteoclasts, we hypothesized that *MPDZ* functionally regulates osteoblast function and impacts bone mass accrual. To test this, we used *Mpdz*-floxed mice to generate primary osteoblast cultures and *in vivo* mouse models that were deficient in *Mpdz* expression. By transducing *Mpdz*-floxed mouse calvarial osteoblasts with adenovirus expressing Cre recombinase, we show that Cre-mediated knockdown of *Mpdz* impairs the formation of mineralized nodules as well as reduces the expression of key osteoblast markers. Furthermore, we used the global Cre driver mouse E2a-Cre to generate mice that were heterozygous for one functional *Mpdz* allele ($Mpdz^{\hat{I}\square/+}$) and performed micro-computed tomography (MicroCT) analysis of skeletal specimens to pinpoint differences in bone microarchitecture. MicroCT reveals a 23% reduction (P Mpdz$^{\hat{I}\square/+}$ mice in comparison to sex-matched wildtype controls. Together, these results demonstrate that *Mpdz* impacts in vitro and *in vivo* bone function and further underscores *MPDZ* as the likely causal gene influencing the BMD association at 9p23.

PrgmNr 3484 - A Novel Role for HAPLN1 in Skeletal Development

View session detail

Author Block: **C. Patel**[1], A. Khanshour[2], Y. Kidane[2], R. Cornelia[2], J. J. Rios[3], C. A. Wise[4]; [1]Univ. of Texas Southwestern, Dallas, TX, [2]Scottish Rite Hosp. for Children, Dallas, TX, [3]Texas Scottish Rite Hosp. for Children, Dallas, TX, [4]Texas Scottish Rite Hosp, Dallas, TX
Disclosure Block: **C. Patel:** None.
Adolescent idiopathic scoliosis (AIS) is a common spine deformity of childhood. Despite its high heritability, few risk-associated loci have been identified. Using a multi-ethnic genome-wide association study meta-analysis, we detected significant association of alleles within genes involved in cartilage development. Here, in a collaboration with the Gabriella Miller Kids First Initiative, we performed whole-genome sequencing of 72 AIS families, which identified rare co-segregating nonsynonymous variants in the *HAPLN1* (p.(Gly336Ser) and p.(Cys304Ser)) gene in two families. *HAPLN1* encodes the Hyaluronan and Proteoglycan Link Protein 1, a protein that binds hyaluronic acid and proteoglycans in the extracellular matrix (ECM) and is thought to stabilize and hydrate cartilage and intervertebral disc. *In vitro* evidence also suggests that an amino-terminal peptide cleaved from HAPLN1 stimulates ECM production. The *HAPLN1* p.(Cys304Ser) variant is predicted to disrupt an evolutionarily-conserved disulfide bridge in the carboxy-terminal loop domain of HAPLN1. To evaluate a possible deleterious effect of the p.(Cys304Ser) allele, we transiently expressed epitope-tagged mutant mouse *Hapln1*^{C306S}, orthologous to the human variant, and a truncated protein lacking the C-terminal loop domain (*Hapln1*^{C306fs*5}) in HEK293T cells. While western blotting detected wild type (WT) Hapln1 protein in the cell lysate, media, and ECM fractions, mutant Hapln1^{C306S} protein was detected only in the lysate and media, and the truncated Hapln1^{C306fs*5} protein was detected only in the cell lysate. Immunofluorescence (IF) microscopy further demonstrated that mutant Hapln1^{C306S} protein was not integrated into the ECM. We engineered mice harboring the *Hapln1*^{C306S} or *Hapln1*^{C306fs*5} alleles using CRISPR/Cas9 gene editing. *Hapln1*$^{C306fs*5/C306fs*5}$ mice were born alive but died shortly after birth, exhibiting a chondrodysplastic phenotype with shortened limbs and dome-shaped head. Although *in vitro* evidence suggests mutant *Hapln1*^{C306S} fails to localize to the ECM, homozygous *Hapln1*$^{C306S/C306S}$ mice were viable, born at expected Mendelian ratios, and showed no significant skeletal abnormalities up to one year of age. Consistent with our *in vitro* results, IF analysis of neonatal spine and long bones failed to detect Hapln1 protein in the growth plates of *Hapln1*$^{C306S/C306S}$ or *Hapln1*$^{C306fs*5/C306fs*5}$ mice, while Hapln1 was readily detected in the hypertrophic zone of the growth plate in control mice. Collectively, our data suggest a novel function for HAPLN1 that is independent of its role in the ECM, and that loss of ECM-associated HAPLN1 may contribute to inherited AIS.

PrgmNr 3485 - A SNP associated with risk for non-syndromic orofacial cleft affects expression of *FOXE1* in an oral epithelium cell line

View session detail

Author Block: **P. Kumari**[1], R. Friedman[2], S. W. Curtis[3], E. J. Leslie[3], M. A. White[2], B. A. Cohen[2], R. A. Cornell[1]; [1]Dept. of Anatomy and Cell Biology, Univ. of Iowa, Iowa city, IA, [2]Dept. of Genetics, Washington Univ. in St. Louis, St. Louis, MO, [3]Dept. of Human Genetics, Emory Univ., Atlanta, GA
Disclosure Block: **P. Kumari:** None.
Genome-wide association studies and linkage analyses have identified two loci near *FOXE1* where single nucleotide polymorphisms (SNPs) in non-coding DNA are associated with risk for non-syndromic orofacial cleft. We predict that a functional subset of such SNPs have allele-specific effects on the activity of oral epithelium enhancers and the rest are merely in linkage disequilibrium with the functional subset. We previously identified a candidate for a functional SNP in one of the loci and here focus on the other. We tested 11 SNPs in this locus for their effects on enhancer activity in a massively parallel reporter assay (MPRA) in the GMSM-K human fetal oral epithelium cell line, and tested a top-scoring SNP from the MPRA in traditional luciferase reporter assays. One SNP, (rs10984103), had allele-specific effects in both MPRA and luciferase assays and lies in chromatin marked as an active enhancer by chromatin marks and nucleosome accessibility. The SNP lies in a predicted binding site for GRHL3 which is interesting because a SNP in an exon of *GRHL3* is associated with risk for non-syndromic cleft palate only. We use CRISPR-mediated homology directed repair to engineer clones of GMSM-K cells to be homozygous for either the risk-associated or non-risk associated alleles of the SNP. In the cells harboring the risk allele relative to those harboring the non-risk allele, expression of *FOXE1* (determined by qRT-PCR), binding of GRHL3 (determined by ChIP-qPCR), and the H3K27Ac signal were all reduced. This study suggests a mechanism for how common variants at *GRHL3* and *FOXE1* affect risk for orofacial cleft. This study also illustrates a generalizable method to screen non-coding SNPs associated with orofacial cleft for those that are functional. This study is funded by NIH DE027362, USA.

PrgmNr 3486 - Clinical application of long-read sequencing technology for unsolved cases of neurodevelopmental disorders

View session detail

Author Block: **S. Hiatt**[1], J. M. J. Lawlor[1], L. H. Handley[1], C. B. Plott[1], E. Partridge[1], L. B. Boston[1], M. Williams[1], J. Jenkins[1], K. M. Bowling[2], J. Grimwood[1], J. Schmutz[1], G. M. Cooper[1]; [1]601 Genome Way, Huntsville, AL, [2]HudsonAlpha Inst. for Biotechnology, Huntsville, AL

Disclosure Block: **S. Hiatt:** None.

Exome and genome sequencing (ES/GS) have proven to be effective tools for the diagnosis of neurodevelopmental disorders (NDDs), many of which are due to highly penetrant, often *de novo,* variation. However, while discovery power and diagnostic yield of genomic testing have consistently improved over time, most NDDs cannot be attributed to currently detectable genetic variation.

One potential approach to overcome variant detection limitations in short-read technology is to use long-read sequencing. We and others have recently assessed long-reads produced using the Pacific Biosciences circular consensus sequence technology, called HiFi, which are both long (>10 kb) and highly accurate (>99% bp accuracy). HiFi reads exhibit better alignment to the reference assembly, leading to more accurate and comprehensive variant calls, including for both SNVs and structural variants (SVs). Furthermore, HiFi reads allow for production of high-quality, phased *de novo* assemblies with N50 values approaching that of hg38. These phased assemblies are critical in understanding complex structural variation, including two of six probands in our initial study.

Building upon the success of our recent pilot project, we are continuing to generate HiFi genomes in probands with unexplained NDDs. These individuals had been previously sequenced with a standard short-read approach, and often been subject to numerous additional clinical tests, yet still lack a molecular diagnosis. We are generating 20-25x CCS coverage for ~8 probands each month, with a current aggregate total of 17 probands. We are using these data to identify disease-causal variants in each proband including SNVs, SVs, and expansions of repeat regions. For each proband, we also create *de novo* assemblies, using short-read data from each parent to facilitate phasing.

As early studies have shown that structural variation may represent a significant source of diagnostic variation in these probands, we are currently developing an in-house database of SVs that will aid in filtration of common variation identified with our HiFi based process. We are anticipating substantial improvements to the analysis of HiFi data and believe this will lead to a major improvement in rare disease diagnosis.

While the initial rates of observation of variants of clinical and research interest (2 of 6, 33% in our cases) in NDD trios were very promising, an extension of this pilot, with over 100 probands sequenced within a year will aid in our understanding of the strengths and diagnostic potential of this approach.

PrgmNr 3487 - Functional interpretation of TAB2 missense variants

View session detail

Author Block: **W. Xu**[1], A. Graves[1], C. Magyar[1], P. E. Bonnen[2], L. Zhang[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Baylor Coll. Med., Houston, TX
Disclosure Block: **W. Xu:** None.
Loss of function mutations in TAB2 (or TAK1 binding protein 2) have been associated with structural heart disease. We reviewed 11 new patients carrying TAB2 null variants from Baylor Genetics Laboratory (BGL) and found a high incidence of broad spectrum of additional developmental abnormalities beyond congenital heart defects, expanding the clinical spectrum of TAB2 haploinsufficiency. Importantly, most missense variants of TAB2 are currently classified as variants of unknown significance in ClinVar. The few reported as pathogenic are also without detailed functional analysis. Thus, we set out to establish a functional assay to assign pathogenicity of missense TAB2 variants. TAB2 has been shown to activate NF-ÎºB and MAPK signaling pathways. We tested the normal allele and additional 47 variants using luciferase reporter assays for NF-ÎºB or AP1 activity in HEK293T cells. These variants were chosen from both BGL and gnomAD database to capture a broad spectrum of variants. Four have been previously reported in the literature as pathogenic (Q230K, P208S, E569K and Q540R), one non-sense variant Q127X, one presumed benign due to high prevalence (H158Y) and 41 variants with unknown significance. We found that NF-ÎºB reporter assay is only able to identify one novel variant (C684Y) as complete loss of function, similar to the non-sense allele and GFP control. All the other alleles show comparable activities. In contrast, the AP1 reporter was able to assign all the known pathogenic variants correctly. Previously reported Q230K, Q127X, C684Y and 5 other variants show complete loss of function. Previously reported P208S shows partial loss of function. We were also able to identify the previously reported E569K and Q540R as gain of function (>2.5 fold). These two variants together with other gain of function variants in TAK1 have previously been reported in patients with autosomal dominant frontometaphyseal dysplasia, a distinct condition from those with TAB2 haploinsufficiency. The normal allele and H158Y show normal AP1 activity. Altogether in the 41 variants of unknown significance, we reclassified 6 complete loss of function, 20 partial loss of function, 10 normal function and 6 partial gain of function. As individuals with partial gain of function (

PrgmNr 3488 - Germline variation in the human immunoglobulin heavy chain locus contributes to inter-individual variability in the expressed antibody repertoire

View session detail

Author Block: **O. Rodriguez**[1], Y. Safanova[2], C. Silver[1], W. Gibson[1], D. Tieri[1], K. Shields[1], J. Kos[1], H. Ke[3], M. Emery[4], G. Deikus[4], R. Sebra[4], K. Jackson[5], S. D. Boyd[6], M. Smith[1], W. Marsco[3], C. T. Watson[1]; [1]Univ. of Louisville, Louisville, KY, [2]Univ. of San Diego, San Diego, CA, [3]Dana-Farber Cancer Inst., Boston, MA, [4]Icahn Sch. of Med. at Mount Sinai, New York, NY, [5]Stanford Univ., Stanford, CA, [6]Stanford Univ Med Ctr, Stanford, CA
Disclosure Block: **O. Rodriguez:** None.
Antibodies (Abs) consist of two identical heavy and light chains and are critical for an effective immune response. The heavy chain is encoded by genes at the immunoglobulin heavy chain (IGH) locus, one of the most structurally complex regions of the human genome, characterized by elevated levels of single nucleotide variants (SNVs) and large structural variants (SVs). There is mounting evidence that polymorphisms within IGH influence inter-individual variation in the expressed Ab repertoire, which may contribute to differences in Ab function in human health and disease. However, the extent to which genetics influences the Ab response has not been characterized. We followed an expression quantitative trait locus framework to identify IGH variants that affect the expressed naive (IgM) and memory (IgG) Ab repertoires in peripheral blood. Using a targeted long read sequencing approach, we generated paired IGH genomic and Ab repertoire sequencing datasets for 166 healthy donors (Asian, n=30; Hispanic, n=20; African American, n=20; Caucasian, n=96). We found that, within the naive IgM repertoire, 2,830 SNVs in IGH were significantly associated (P-5) with variation in usage for 66% (29/44) of commonly used (>0.001% of repertoire) variable (IGHV) genes. Identified variants included both coding and non-coding SNVs, as well as six common SVs and rare deletions spanning up to 19 genes. We found variants were either associated with single IGHV genes, or multiple genes, in some cases revealing complex long-range effects. For example, we identified an intergenic SNV associated with the usage of five genes (*IGHV3-33*, *IGHV3-53*, *IGHV4-59*, *IGHV3-66*, and *IGHV1-69*) that span a distance of >400Kb. Interestingly, the SNV reference allele frequency differed between populations (African, 0.78; European, 0.34). We also observed complex interactions between variants located within multiple regions of the locus. For example, the usage of IGHV4-38-2 was impacted by a SV and multiple SNPs separated by ~900 Kb. Finally, 85% (2,426/2,830) of significant IgM SNVs were also associated (P-5) with IGHV usage in the IgG repertoire, indicating that genetic factors influencing V(D)J recombination also have downstream impacts on the repertoire following antigen stimulation. Our study is the first to comprehensively characterize links between IGH polymorphism and inter-individual variation in the Ab repertoire, revealing an underappreciated role for IGH variants in the immune response. These findings set a foundation for investigating the molecular mechanisms influencing Ab repertoire development, and will ultimately provide insight into Ab-mediated immunity in disease.

PrgmNr 3489 - High Throughput Trafficking Assay Identifies Novel Loss and Gain of Function Variants in the Potassium Channel Gene *KCNE1*

View session detail

Author Block: **A. Muhammad**[1,2], T. Yang[2], M. A. Blair[2], L. D. Hall[2], K. Matreyek[3], D. M. Fowler[4], D. M. Roden[2], A. M. Glazer[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [4]Univ of Washington, Seattle, WA
Disclosure Block: **A. Muhammad:** None.
Introduction: *KCNE1* encodes a 129-amino acid transmembrane subunit of I_{Ks}, a potassium current important for cardiac repolarization. Loss-of-function variants in *KCNE1* can cause long QT syndrome, associated with fatal arrhythmias. Clinical testing has identified large numbers of variants of uncertain significance (VUSâ□□s) in *KCNE1*, necessitating novel approaches to accurately classify variants at scale.
Objective: We conducted a massively parallel cell surface abundance assay on a library of nearly all possible amino acid variants in *KCNE1*.
Methods: Using degenerate primers at each codon, we used inverse PCR to create a comprehensive library of variants in *KCNE1* with an extracellular HA tag. The library was expressed in a one-variant-per-cell format using a Bxb1 landing pad in HEK293 cells engineered to stably express *KCNQ1*, the other subunit of I_{Ks}. Cells were stained with a fluorescent anti-HA antibody and sorted into 4 equally populated groups of increasing surface abundance using flow cytometry. High throughput sequencing of barcode DNA from cells in each group was used to quantify the frequency of variants across the four groups, and surface abundance of KCNE1 was quantified using a weighted average, with possible values from 0 to 3. WT-like trafficking was defined as scores within the 95% confidence interval of the mean for synonymous variants.
Results: The library represented 80.1% of all possible missense variants in KCNE1. Surface abundance scores of these variants were normally distributed (1.53Â±0.43). Previously published trafficking-deficient variants had low scores in the assay. Most mutations at amino acids 35-38 reduced KCNE1 surface abundance, whereas positions 47-59 in the transmembrane helix had decreased surface abundance when mutated to most polar or charged amino acids. Surface abundance also decreased for aromatic or larger aliphatic amino acids introduced in the cytoplasmic domain at positions 79-114 which some reports implicate as a binding site for KCNQ1. We also identified variants across the protein with a novel, gain-of-function phenotype, increasing I_{Ks} at the cell surface. Nonsense variants before amino acid 57 in KCNE1 reduced surface expression, but nonsense variants after amino acid 57 had WT-like scores. Of variants currently classified as VUS, 23% had a trafficking score greater or less than WT-like.
Conclusion: Using deep mutational scanning, we interrogated ~2,000 variants in a clinically important ion-channel subunit, KCNE1. Our data provide structural insights and functional information to facilitate interpretation of VUSâ□□s, and uncover a gain-of-function phenotype suggesting a novel arrhythmia mechanism.

PrgmNr 3491 - Hyperproliferative lymphatic sprouting caused by an NRAS mutation can be rescued by targeting PI3K/AKT and RAS/MAPK signaling pathways in cell-based modeling systems

View session detail

Author Block: **M. Battig**[1], M. March[1], L. Matsuoka[1], S. Sheppard[1], C. Kao[1], C. Seiler[2], D. Li[1], H. Hakonarson[1]; [1]Ctr. for Applied Genomics, Children's Hosp. of Philadelphia, Philadelphia, PA, [2]Aquatic Zebrafish Core, Children's Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **M. Battig:** None.
Generalized lymphatic anomaly (GLA) and its more aggressive subtype kaposiform lymphatic anomaly (KLA) are intractable lymphatic malformations that are associated with severe clinical symptoms and poor prognoses. These lymphatic malformations occur primarily in bones, viscera, and thoracic and abdominal cavities, and will lead to complications such as organ compromise, pleural and pericardial effusions, and ascites. In patients with KLA, effusions and ascites are often hemorrhagic. The etiologies of these diseases are not fully understood, though studies have begun to identify causal genetic variations. In particular, a somatic mutation in *NRAS* (c.182A>G:p.Q61R) was recently found in tissue samples of GLA and KLA patients. *NRAS* encodes the GTPase N-ras, which serves an important role in lymphangiogenesis by regulating cell proliferation. To understand how this mutation may affect lymphangiogenesis, we expressed the mutation in human dermal lymphatic endothelial cells (HDLEC) and characterized their biochemical and functional activity. Immunoblotting of PI3K/AKT and RAS/MAPK signaling cascade components revealed that p.Q61R leads to increased activity in both pathways. Moreover, the activity could be restored using the MEK inhibitor trametinib and the mTOR inhibitors rapamycin (mTORC1 inhibitor) and OSI-027 (mTORC1 and mTORC2 inhibitor). In lymphatic sprouting assays, spheroids of HDLEC expressing NRAS(p.Q61R) exhibited a significantly greater number of capillary-like structures in comparison to the wild-type control. This hyperproliferative phenotype could be rescued by using either OSI-027 or trametinib, as seen by a reduction in the number of sprouts emanating from the spheroid and their length. Moreover, OSI-027 and trametinib, when applied together, inhibited increased sprouting at lower concentrations. These findings indicate that complex lymphatic anomalies may arise from mutations activating multiple signaling pathways involved in the growth of lymphatic endothelial cells, and that inhibitor-based therapies may benefit patients with these diseases, particularly if multiple signaling pathways are being targeted.

PrgmNr 3492 - Seeking answers for ocular disease: Characterizing intronic variants of uncertain significance using a minigene system

View session detail

Author Block: **C. Leavens**[1], K. Sadowska[1], D. Reither[1], S. Schuetze[2], G. J. Fischer[2], A. Gruber[2], D. Gingerich[1], J. Lyman Gingerich[1]; [1]Univ. of Wisconsin Eau Claire, Eau Claire, WI, [2]PreventionGenetics, Marshfield, WI

Disclosure Block: **C. Leavens:** None.

Exome and genome sequencing is quickly becoming standard healthcare practice for many rare diseases; however, effectively interpreting genetic variants remains a challenge. Variants of uncertain significance (VUS) are frequently included on clinical sequencing reports, however their clinical relevance is unclear, posing a challenge for clinicians and patients seeking a molecular diagnosis. In most cases, VUS require additional family studies or functional assays to resolve the uncertainty. One large group of VUS are intronic variants predicted by *in silico* analyses to potentially affect pre-mRNA splicing. In partnership with a clinical laboratory, a cursory database analysis identified nearly 1,000 reported VUS which may impact splicing and require a functional assessment. Our pilot analysis focused on 30 variants with the strongest *in silico* splice disruption predictions in genes associated with ocular diseases. A well-characterized minigene system was used to assess the predicted splice defect (PMID: 28679633; 16925019). Briefly, the minigene system involves cloning a single gene segment into a plasmid vector which is then transfected into eukaryotic cells. Processed mRNA transcripts are then sequenced to determine the effects of the variant on splicing. To date, we have completed transfection of nearly a third of the variants, with a potential splicing impact observed in at least 5 variants. We will present our minigene assay findings in comparison to the *in silico* predictions. Ultimately, our findings may impact variant interpretation and the clinical diagnosis for each patient.

PrgmNr 3493 - Single-cell RNA sequencing of red and black melanocytes uncovers new pathways in pigmentation and disease

View session detail

Author Block: **H. Berns**, D. Watkins-Chow, W. Pavan; NIH, Bethesda, MD
Disclosure Block: **H. Berns:** None.
Melanoma is the most lethal form of cutaneous cancer, and variants in the melanocortin-1-receptor (MC1R) are associated with greater than 60% increased disease risk. Some human MC1R loss-of-function/hypomorphic variants produce disproportionately high levels of red-yellow pigment (pheomelanin), and individuals carrying these variants present with fair skin, poor tanning response, and increased susceptibly to melanoma development. Activation of photoprotective black-brown pigment (eumelanin) production downstream of MC1R is well established, however, there is still much to be learned regarding global gene expression changes upon activation or inhibition of MC1R. Thus, we aimed to characterize gene expression profiles of primary melanocytes isolated from mouse models with differing levels of MC1R activation by comparing red melanocytes, from a Lethal Yellow (Ay/a) mouse, with black melanocytes, from a Non-agouti (a/a) mouse. Using a single-cell RNA-sequencing approach, we identified 568 differentially expressed genes between these two primary melanocyte populations (FDR

PrgmNr 3494 - A computational approach for detecting physiological homogeneity in the midst of genetic heterogeneity

View session detail

Author Block: **P. Zhang**; The Rockefeller Univ., New York, NY
Disclosure Block: **P. Zhang:** None.
The human genetic dissection of clinical phenotypes is complicated by genetic heterogeneity. Gene burden approaches that detect genetic signals in case-control studies are underpowered in genetically heterogeneous cohorts. We therefore developed a genome-wide computational method, network-based heterogeneity clustering (NHC), to detect physiological homogeneity in the midst of genetic heterogeneity. Simulation studies showed our method to be capable of systematically converging genes in biological proximity on the background biological interaction network, and capturing gene clusters harboring presumably deleterious variants, in an efficient and unbiased manner. We applied NHC to whole-exome sequencing data from a cohort of 122 individuals with herpes simplex encephalitis (HSE), including 13 cases with previously published monogenic inborn errors of TLR3-dependent IFN-Î±/Î² immunity. The top gene cluster identified by our approach successfully detected and prioritized all causal variants of five TLR3 pathway genes in the 13 previously reported cases. This approach also suggested candidate variants of three reported genes and four candidate genes from the same pathway in another 10 previously unstudied cases. TLR3 responsiveness was impaired in dermal fibroblasts from four of the five cases tested, suggesting that the variants detected were causal for HSE. NHC is, therefore, an effective and unbiased approach for unraveling genetic heterogeneity by detecting physiological homogeneity.

PrgmNr 3495 - A statistical analysis framework to test the presence of allelic imbalance from a genomic region using allelic count data

View session detail

Author Block: **J. Choi**, D. M. Brown, B. M. Wilk, M. Gajapathy, E. Worthey; Ctr. for Computational Genomics and Data Sci. (CGDS) Dept. of Pediatrics, UAB Sch. of Med., Univ. of Alabama at Birmingham, Birmingham, AL
Disclosure Block: **J. Choi:** None.
Mosaicism is a biological phenomenon characterized by coexistence of genetically distinct cell populations in an individual developed from a single zygote. Diverse forms of mosaicism have been recognized from individuals with developmental disorders and cancer. For example, developmental disorders can be seen from individuals with mixploidy where diploid cells coexist with triploid cells due to prezygotic or postzygotic errors. On the other hand, in cancer, accumulation of somatic mutations can result in mixture of normal diploid cells with cancer cells, which exhibit arbitrary degree of chromosomal abnormalities and ploidy levels.One way to identify mosaicism in an individual is through examination of existence of allelic imbalance in affected tissue. Besides molecular cytogenetic analysis by fluorescence in situ hybridization, single-nucleotide microarrays and next-generation sequencing (NGS) techniques have been used to detect genomic regions with allelic imbalance arising from various forms of mosaicism. For instance, the previous research [1] has demonstrated that the pattern of alternative allele fractions across genomic region from NGS data can be used to identify individuals with mixploidy and to estimate the mixture proportion of diploid and triploid cells, which was shown comparable to cytogenic analysis.
With the motivation to extend the previous research [1], we developed a statistical analysis framework to detect the presence of systematic deviation of alternative allele fractions in a genomic region from an individualâ□□s genomic data with respect to those from a set of reference genomic data for which, no known abnormality is reported. To test the feasibility, we simulated allelic count data for a set of reference genomic data as well as that for individuals carrying a broader and subtler forms of mosaicism where in addition to a majority of normal cells, a minority of abnormal cells is assumed to have a variable allelic count ratio that leads to a shift in allelic balance. We showed that the statistical analysis framework based on a regression analysis technique applicable for a count data can achieve 80% of power to detect the presence of about 5 - 10% abnormal cells carrying the majority of possible allelic imbalance allowed for a given ploidy level, when genomic loci subject to the allelic imbalance span sufficiently large region and the over-dispersion of the count data is small enough. We are currently investigating ways to improve the detection power based on the current framework and to determine utility when coupled with low coverage whole genome sequencing.
[1] Holt, James M., et al, BioRxiv 2018

PrgmNr 3496 - Assessing Short-Read Utility for SVs

View session detail

Author Block: **A. ENGLISH**[1,2], V. Menon[1], R. Flickinger[2], G. A. Metcalf[1], R. A. Gibbs[1], F. J. Sedlazeck[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Spiral Genetics, Seattle, WA
Disclosure Block: **A. English:** Salary/Employment; Spiral Genetics.
Structural Variation (SV) is a major source of genomic variation and the cause of multiple genomic diseases. SVs (insertions/deletions larger than 50bp) are difficult to identify within Illumina short-read data, prompting the use of long-read sequencing. Despite the overall success of long-read methods for detecting SVs, the data generation costs and sample requirements are limiting. Improved characterization of the performance of short-read sequencing for SV detection - including a detailed comparison with long-read methods, the impact of different genome references, and the source of sample data - can therefore enhance routine SV analyses. We leveraged previously published long-read, haplotype-resolved Human genome assemblies to create high-confidence sets of SVs (hcSVs) from 36 individuals from 5 ethnicities, employing 3 different references (Hg19; GRCh38; CHM13). Using 156 short-read samples of the same individuals from 2 previously published projects, we assessed the performance of two SV discovery methods (BioGraph, Manta) across the hcSVs. We also tested the â□□genotypabilityâ□□ of hcSVs using short-reads by leveraging two graph-reference based genotypers (BioGraph, Paragraph). We found that the average Human genome contains 26,672 SVs (10k deletions, 16k insertions) compared to GRCh38 and 24,525 SVs (12k deletions, 12k insertions) compared to CHM13. We note an increase in the SV counts in samples of African ancestry (~29k GRCh38 SVs) compared to those of non-African ancestry (~25k GRCh38 SVs). The short-read methods are able to discover ~70% of hcSVs and accurately genotype ~80%, highlighting that the underlying signals for SVs are conserved across sequencing technologies. However, for SV discovery we observe a bias against mid to large-sized insertions. Recovery improves for SV genotyping, but relies on an initial list of known hcSVs over which to genotype. We find that ~10% of hcSVs occur near or across genes, and of these, ~90% are accurately identifiable by short-read sequencing. This work highlights that while short-read sequencing captures fewer SVs than long-reads, much of the signal elucidated by both approaches may still allow for a better understanding of SVs and their geno- and phenotypic implications when appropriate tools and analytical methods are employed.

PrgmNr 3497 - Copy-Number Variation of CTCF Binding Sites Associates with Diverse Clinical Phenotypes

View session detail

Author Block: **C. Tubbs**[1], L. Han[2], M. Benton[3], E. McArthur[1], J. Capra[4], D. Ruderfer[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Baylor Univ., Waco, TX, [4]Univ. of California at San Francisco, San Francisco, CA
Disclosure Block: **C. Tubbs:** None.
Background: CCCTC-binding factor (CTCF) is an integral component in the formation of chromatin contact domains that underlie 3D genome structure. There is increasingly strong evidence for the utility of this organization in the proper regulation of gene expression. Recently, we have observed signatures of purifying selection acting against copy-number variants (CNVs) that disrupt CTCF binding sites. These events hold the potential to dysregulate the expression of many genes through the alteration of 3D genome organization. Gene misexpression driven by this mechanism has been causally linked to a variety of rare limb malformities and, in several cases, been shown to contribute to the pathology of developmental disorders and cancers. However, we lack a comprehensive understanding of when and how CTCF disruption by CNVs contributes to variation in common traits and disease. **Methods:** We address this gap by quantifying the relationship between CNVs, CTCF binding sites, and human traits using 56,112 genotyped individuals of European descent in BioVU, Vanderbiltâ□□s biobank linked with electronic-health records (EHR). We identified 398,131 rare (AF 20kb) CNVs, consisting of 175,475 deletions and 222,656 duplications. Individuals in our cohort, on average, carried 3.74 CNVs that covered 447kb of sequence. To determine how these CNVs impact CTCF binding sites, we curated CHIP signal from ENCODE across 89 human tissues to yield a final genome-wide set of 153,535 sites. Finally, to understand the relationship between CNV disruption of CTCF and human traits, we defined 1,865 clinical phenotypes using diagnostic billing codes ascertained from EHR data (â□□phecodesâ□□). **Results:** We hypothesize that variation in CTCF is deleterious and contributes to common disease risk through changes in chromatin topology. We assessed the association between CNVs that overlap CTCF binding sites with the clinical phenome defined by phecodes using a logistic regression model adapted from PLINKâ□□s CNV-enrichment test. We found significant associations between CNVs that disrupt CTCF binding sites and 79 phecodes across 13 clinical phenotype domains, all of which survived multiple testing correction (Bonferroni, p Conclusion: Our findings reveal a diverse array of clinical phenotypes that are enriched for CNVs that have overlap with CTCF binding sites. Ongoing work will aim to isolate and disentangle the direct contributions of CNVs altering CTCF to these traits.

PrgmNr 3498 - Evaluation of long-reads across challenging medical relevant genes and its implications for All of Us

View session detail

Author Block: **M. Mahmoud**[1], K. Garimella[2], P. Audano[3], S. Huang[2], N. Prasad[4], R. E. Handsaker[2], S. Hall[4], A. Pionzio[4], M. E. Talkowski[5], E. E. Eichler[6], S. E. Levy[7], F. J. Sedlazeck[8]; [1]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [2]Broad Inst. of MIT and Harvard, Cambridge, MA, [3]The Jackson Lab., Farmington, CT, [4]HudsonAlpha Discovery, Owens Cross Roads, AL, [5]Massachusetts Gen. Hosp., Boston, MA, [6]Univ of Washington, Seattle, WA, [7]Hudson Alpha Inst. for Biotechnology, Huntsville, AL, [8]Baylor Coll. Med., Houston, TX

Disclosure Block: **M. Mahmoud:** None.

The overall heritability explained by genetic studies to date is low for most complex diseases. Several common variant studies have suggested associations with loci in proximity to complex genomic regions that, in aggregate, harbor ~400 relevant medical genes (eg. *SMN1*, *RHCE*, *NCF1*, *LPA*, *TPO*, *TNNT1* and *HLA*), many of which are refractory to alignment using short-read genome sequencing. The emergence of long-read technologies hold great promise for detection and phasing of short (SNVs) and structural variants (SVs) to disentangle the complex interactions of these mutations. Unfortunately, the error rate and high costs limit availability and application in clinical sequencing. The All of Us Research Program long-read project aims to widen our understanding of the sequence diversity in complex genomic regions that potentially have direct implications on discovery of variation in medically relevant genes. We have previously identified 395 genes that are inaccessible using short-reads, which are also found in genetic panels and are highlighted in multiple studies citing their medical importance. Moreover, mutations in these genes cause different diseases such as Cancer, spinal muscular atrophy, and Rh Deficiency Syndrome. Using these genes, we investigate the benefits and limitations of long-reads across 27 samples including 4 trios using Illumina, PacBio HiFi and Oxford Nanopore Technology (ONT). Each sample was sequenced with one flow cell, enabling a systematic comparison across the technologies. Assessing the boundaries of ONT and PacBio HiFi across these 395 medical relevant genes revealed that we can haplotype-resolve SVs and SNVs across a majority (80%) of these genes, including *LPA*, with a complete loss of coverage in 3 autosomal genes (eg. *CORO1A*) and 14 genes on chrY (eg. *CDY1*). A few of these genes are interestingly sample-dependent, thus underlying their complexity compared to a reference issue. We identified a slight advantage for ONT attributable to their longer read length for mapping to repeats, such as HLA, but exhibit a higher indel error rate compared to PacBio HiFi. For at least 85.50% of genes, we were able to phase SNV and SV to study the trans vs. cis relationships of coding and noncoding effects. This study presents the strength and limitations of long-reads and their potential role to improve clinical WGS sequencing. Although cost and availability are still significant limitations, long reads can improve SNV and SV detection in clinical settings. These findings have clear implications for All of Us, human adaptation, and translational research. In our presentation, we will highlight these insights and share our progress in the All of Us program.

PrgmNr 3499 - Framework and analysis of cell-specific expression contributions to Mendelian Disease

View session detail

Author Block: **J. Whitlock**[1], T. Soelter[2], K. McPherson[2], T. Howton[1], V. Oza[2], B. N. Lasseigne[2]; [1]Univ. of Alabama at Birmingham, SOM, birmingham, AL, [2]Univ. of Alabama at Birmingham, SOM, Birmingham, AL

Disclosure Block: **J. Whitlock:** None.

There are approximately 10,000 Mendelian diseases affecting 25 million Americans. Mendelian diseases are caused by germline aberrations that are present in all cells of the body but typically manifest in a number of tissues. Despite advances in genomic sequencing, gaps remain in translating individual genomic variation to observed phenotypic outcomes in rare mendelian disease. As a result, nearly a third of Mendelian diseases have unknown molecular bases. Moreover, for the remaining two-thirds, how molecular variation leads to disease is poorly understood in one-third of cases. There are multiple hypothesized drivers of tissue-specificity such as gene expression, regulatory networks, and non-cell-autonomous based mechanisms. Additionally, studies have proposed some diseases are not driven solely by tissue-specific expression, but rather underlying cell-specific disease manifestations. These remain unidentified or masked in bulk RNA-sequencing data. Many of these mechanisms remain largely uninvestigated across multiple cell or tissue types in a normal state. With this, comes a lack of understanding of the impact patient-specific variants may have on altering healthy patterns and leading to disease. A handful of tissue-specific gene expression tools exist. However, there is only one method for calculating cell-specific expression of genes, and it lacks a package for widespread application. Here, using renal and neurological associated human Mendelian disease genes in OMIM as a test set, we assess cell-type contributions to bulk RNA-sequencing data and develop a framework for assessing mechanisms behind cell-specific gene expression.

PrgmNr 3500 - Haplotype resolved LPA and Kringle repeats to study minorities across the USA

View session detail

Author Block: **F. J. Sedlazeck**[1], M. Mahmoud[1], V. Menon[2], B. Yu[3], H. Chen[4], S. N. Jhangiani[1], Q. Meng[5], Y. Han[1], H. Doddapaneni[1], D. M. Muzny[1], R. A. Gibbs[1], D. R. Murdock[5], C. Ballantyne[1], G. A. Metcalf[1], E. Boerwinkle[3]; [1]Baylor Coll. Med., Houston, TX, [23]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [4]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [5]Baylor Coll. of Med., Houston, TX

Disclosure Block: **F.J. Sedlazeck:** None.

Heart disease is the number one cause of death in the United States and many risk factors are highly heritable, including levels of Lp(a) lipoprotein. Circulating levels of Lp(a) have been associated with risk of coronary heart disease (CHD) and stroke. Apolipoprotein(a), a precursor to Lp(a), is encoded by the LPA gene, and LPA includes multiple copies of the Kringle repeats (IV) that are negatively correlated with its expression. The analysis of LPA and the kringle IV repeats is often hindered with short reads due to its repetitive and complex nature, however previous studies have identified two single nucleotide polymorphisms that are strongly associated with an increased level of Lp(a), a reduced number of Kringle repeats, and an increased risk of cardiovascular disease. However, this association isnâ□□t well replicated in non-caucasian ethnicities, so the utility is limited when using these variants to assess the genetic risk of cardiovascular disease in certain populations. We found this to be true in our clinical cardiovascular genomics project, HeartCare. In this study, despite the presence or absence of the indicative SNPs, several non-Caucasian participants had unpredictable Lp(a) protein levels overall, highlighting that the linkage of the SNP and the kringle IV repeat numbers are not conserved across all ethnicities and further study of the LPA gene is needed to establish accurate genetic risk profiles for cardiovascular disease in diverse populations. We selected 20 Caucasians (10) and Hispanics (10) individuals that have extremely high or low levels of Lp(a) protein levels for long-read Oxford Nanopore sequencing. We were able to call SNVs and the tandem duplications of the kringle IV repeats together with other smaller Structural Variants (SVs) present in several participants. We were able to haplotype resolve the entire locus enabling investigation of candidate SNVs in linkage disequilibrium with the number of kringle IV repeats and the amount of linkage disequilibrium present in Hispanics. Using these genomic findings, we identified SNVs and copy number calls in a very large sample of Hispanics and non-Hispanic whites (N=14,111) and relate these variations to plasma Lp(a) levels. These data provide valuable insight into the genetic architecture of plasma Lp(a) levels across diverse populations.

PrgmNr 3501 - Identification of allele-specific KIV-2 repeats among multi-ethnic groups and association with LP(a) measurements

View session detail

Author Block: **S. Behera**[1], X. Chen[2], L. S. Kanikkannan[3], V. K. Menon[1], G. A. Metcalf[1], B. Yu[3], M. A. Eberle[2], E. Boerwinkle[3,1], C. M. Ballantyne[4], R. Kaplan[5], C. J. Rodriguez[5], F. J. Sedlazeck[1]; [1]Human Genome Sequencing Ctr., Baylor Coll. of Med., Houston, TX, [2]Illumina Inc., San Diego, CA, [3]Sch. of Publ. Hlth., Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [4]Dept. of Med., Baylor Coll. of Med., Houston, TX, [5]Dept. of Epidemiology and Population Hlth., Albert Einstein Coll. of Med., Bronx, NY
Disclosure Block: **S. Behera:** None.
Studies on the human *LPA* gene have found evidence that the kringle IV type 2 (KIV-2) variable number tandem repeats (VNTR) variant is one of the controlling factors of lipoprotein(a) [Lp(a)] isoform size. The *LPA* gene including the KIV-2 variant determines the Lp(a) level (a high number of KIV-2 repeats is associated with low Lp(a) concentration) and has strong associations with cardiovascular diseases. Nevertheless, it remains challenging to determine the number of KIV-2 repeats in whole-genome sequencing (WGS) data due to the repetitiveness of KIV-2. Lp(a) is currently widely studied among Europeans with recent papers reporting clear associations with cardiovascular risk. However, it remains challenging to extend these insights to other ethnicities including Hispanics due to the lack of genetic and phenotypic data available on non-Europeans. Thus, an allele-specific copy number (CN) estimation of KIV-2 is needed to improve the genetic diagnosis and understanding of the impact of KIV-2 on cardiovascular risk across ethnicities.

Using data from different cohort studies, we aim to study the association of KIV-2 repeats with the Lp(a) concentrations and cardiovascular risk prediction. To achieve this, we developed a novel approach to directly assess KIV-2 levels derived from the Illumina WGS dataset. We benchmarked this method carefully against Pacbio HiFi based assemblies to ensure high accuracy and precision. We used a WGS dataset of randomly selected 3,020 participants (samples sequenced on Illumina HiSeq X and mapped to GRCh38 reference sequence) from multiple ethnic groups including 1000 European samples, 1019 African-American samples from the Atherosclerosis Risk in Communities (ARIC) cohort study, and 1001 Hispanic samples from the Hispanic Community Health Study and the Study of Latinos (HCHS/SOL).

Our tool estimated the summed copy number of both alleles in all samples and performed haplotype phasing of ~46% of the samples (45.9% Europeans, 51.3% African-Americans, and 40.5% Hispanics). The frequency distribution of CN estimates among three ethnic groups shows that the African-American group has a higher percentage (~70%) of samples that are in KIV-2 repeats ranging from 20 to 40 versus ~45% for the Hispanic group. Using these newly-derived KIV-2 CN estimates, we will present the results of our association study, which utilizes protein measurements as well as health records from each of these individuals. We will present details about the method and the differences we were able to identify across the different ethnicities. Especially the latter has the potential to improve diagnosis of cardiovascular disease risks among understudied ethnicities.

PrgmNr 3502 - Knowledge based artificial intelligence for variant pathogenicity prediction for Mendelian disorders

View session detail

Author Block: **D. Mao**[1], R. Al-Ouran[1], C. Liu[1], L. Wang[1], C. Deisseroth[1], S. Kim[1], L. Li[1], P. Liu[1], B. Yuan[2], S. Yamamoto[1], M. F. Wangler[3], B. Lee[3], Undiagnosed Diseases Network, Baylor Genetics, H. J. Bellen[1], Z. Liu[4]; [1]Baylor Coll. of Med., Houston, TX, [2]Seattle Children's Hosp., Seattle, WA, [3]Baylor Coll. Med., Houston, TX, [4]Duncan Neurological Res. Inst., Baylor Coll. of Med., Houston, TX
Disclosure Block: **D. Mao:** None.
Every year thousands of patients, with potential rare genetic disorders, face uncertainty when healthcare providers are unable to discover the cause for their symptoms. The Undiagnosed Diseases Network (UDN) seeks to provide answers for patients and families affected by these mysterious conditions. The process of defining pathogenicity currently requires labor-intensive manual searches of a variety of databases and web resources. This manual process is time-consuming, subject to inter-user variability and variations in the depth or quality of the databases. It also requires broad expertise across multiple biological and informatics domains. Here, we created a systematic, comprehensive search engine, MARRVEL (Model organism Aggregated Resources for Rare Variant ExpLoration, http://marrvel.org), that mines all the critical information for variant analysis and presents it in a succinct, user-friendly way. MARRVEL integrates human databases (OMIM, gnomAD, ExAC, ClinVar, Geno2MP, DGV, and DECIPHER) and seven model organism databases from yeast to mammals. Furthermore, we are also developing a Knowledge-based Artificial Intelligent system (MARRVEL-AI) to prioritize and identify novel disease-causing coding variants. The interpretability of a machine learning method inversely correlates with its accuracy for complex tasks. To circumvent this, we are combining different models of artificial intelligence with complementary strengths, such as expert system and random forest. With only a small training data set, our model achieved high accuracy in identifying disease-causing variants for UDN cases.

PrgmNr 3503 - MetaRNN: Differentiating rare pathogenic and rare benign missense SNVs and InDels using deep learning

View session detail

Author Block: **C. Li**[1], D. Zhi[2], K. Wang[3], X. Liu[4]; [1]Coll. of Publ. Hlth., Univ. of South Florida, Tampa, FL, [2]Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [3]Children's Hosp. of Philadelphia & Perelman Sch. of Med., Univ. of Pennsylvania, Philadelphia, PA, [4]Univ. of South Florida, Tampa, FL
Disclosure Block: **C. Li:** None.
Separating rare pathogenic and rare benign variants from a large list of candidate mutations in next-generation sequencing experiments is an essential task in exome-sequencing-based Mendelian disease studies. This task is especially challenging for mutations that can cause changes in amino acid (AA) sequences, namely, nonsynonymous single nucleotide mutations (nsSNVs) and non-frameshift insertion/deletions (nfINDELs) while being exempt from definite and severe consequences. In this study, we present the pathogenicity prediction models MetaRNN and MetaRNN-indel to help identify and prioritize these two types of mutations using deep learning and context annotations. Twenty-eight high-level annotation scores and allele frequency features were used to represent the properties of each mutation and a recurrent neural network was adopted to extract feature information from a +/- 1 codon window around the affected codon. Trained using the ClinVar database, the MetaRNN and MetaRNN-indel models outperformed state-of-the-art competitors across all test datasets. Additionally, MetaRNN and MetaRNN-indel scores are comparable and well-calibrated to be combined for an integrated (nsSNV+nfINDELs) rare-variant burden test for genotype-phenotype association. MetaRNN executables and pre-computed scores are available at http://www.liulab.science/MetaRNN.

PrgmNr 3504 - Participation of rs9138 and rs1126616 of *SPP1* gene in gender-specific predisposition in the development of systemic lupus erythematosus

View session detail

Author Block: **A. Rivera-Cameras**[1,2], A. Ruiz-RamÃrez[1,2], I. DÃivalos- RodrÃguez[1,2]; [1]DivisiÃ³n de GenÃ©tica, Centro de InvestigaciÃ³n BiomÃ©dica de Occidente, IMSS., Guadalajara, Mexico, [2]Doctorado en GenÃ©tica Humana, CUCS, Univ. de Guadalajara, Guadalajara, Mexico
Disclosure Block: **A. Rivera-Cameras:** None.
INTRODUCTION: Systemic lupus erythematosus (SLE) is a condition of the immune system. Itâ□□s characterized by exaggerated B-cell and T-cell response, loss of immunity, production and removal of defective antibodies. Sex hormones and environmental influences can contribute to genetically predisposed immune system dysfunction. **OBJETIVE:** Analyze the participation of the variants rs9138 and rs1126616 in gender-specific predisposition in systemic lupus erythematosus. **METHODOLOGY:** The eQTL calculation of the variants was performed in 54 tissues reported in GTEx (V8), were selected as true eQTL when they presented an m-value >0.9. A variant-disease association (VDA) was made in DisGenet platform, based on a VDA score that ranges from 0 to 1 according to how much the association is supported in the different sources (UNIPROT, CLINVAR, GWASCAT, GWASDB). **RESULTS**: It was found that both variants presented an m-value >0.9 only in the testis tissue, with a marked difference in all other tissues, since the rest remained below 0.1. In the variant-disease association, the three diseases with the highest scores for each variant are mentioned, rs1126616 was associated with colorectal carcinoma, urolithiasis, and systemic lupus erythematosus (SLE). rs9138 had an association with rheumatoid arthritis, colorectal cancer, and neoplasms. No statistically significant difference was found between the mean values of the VDA score for the variants (p=0.565, N=20). **DISCUSSION:** The positive association between testicular function and the development of SLE has been previously seen, particularly when there is testicular hypofunction.[1] The key role of the rs9138 variant, in triggering high levels of IFN-Î± and secreted phosphoprotein 1 (SPP1)[2], as well as the remarkable role that IFN-Î± levels play in the development of rheumatological diseases has been seen, many research groups have even proposed that the increase of this cytokine is sufficient to induce SLE.[3] In our opinion, the variant rs1126616 has not been proposed before in the gender-specific predisposition for SLE, but this variant has been correlated as a risk factor for lupus nephritis.[4] The results obtained from the association of the variants with the tissues are interesting for the positive association between testicular hypofunction and SLE, support the hypothesis that low testosterone levels can influence the development of male SLE. It suggests that men with SLE have a higher risk of comorbid, this may justify its consideration in the treatment of patients, this information opens up avenues for future research focused on different tissues.

PrgmNr 3505 - Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data

View session detail

Author Block: **M. Zhang**[1,2], K. Hou[3], K. K. Dey[1,2], K. A. Jagadeesh[1,2], A. O. Pisco[4], M. Gandal[5,6,7], J. Y. Zou[4,8,9], S. Raychaudhuri[10,11,2,12,13], B. Pasaniuc[3,6,14,15], A. L. Price[1,2,16]; [1]Dept. of Epidemiology, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]Program in Med. and Population Genetics, Broad Inst. of MIT and Harvard, Cambridge, MA, [3]Dept. of Pathology and Lab. Med., David Geffen Sch. of Med., Univ. of California, Los Angeles, Los Angeles, CA, [4]Chan Zuckerberg Biohub, San Francisco, CA, [5]Dept. of Psychiatry, David Geffen Sch. of Med., Univ. of California, Los Angeles, Los Angeles, CA, [6]Dept. of Human Genetics, David Geffen Sch. of Med., Univ. of California, Los Angeles, Los Angeles, CA, [7]Program in Neurobehavioral Genetics, Semel Inst., David Geffen Sch. of Med., Univ. of California, Los Angeles, Los Angeles, CA, [8]Dept. of Electrical Engineering, Stanford Univ., Palo Alto, CA, [9]Dept. of BioMed. Data Sci., Stanford Univ., Palo Alto, CA, [10]Ctr. for Data Sci., Brigham and Womenâ□□s Hosp. and Harvard Med. Sch., Boston, MA, [11]Div. of Rheumatology, Inflammation, and Immunity, Dept. of Med., Brigham and Womenâ□□s Hosp. and Harvard Med. Sch., Boston, MA, [12]Dept. of BioMed. Informatics, Harvard Med. Sch., Boston, MA, [13]Ctr. for Genetics and Genomics Versus Arthritis, Manchester Academic Hlth.Sci. Ctr., Univ. of Manchester, Manchester, United Kingdom, [14]Bioinformatics InterDept.al Program, Univ. of California, Los Angeles, Los Angeles, CA, [15]Dept. of Computational Med., David Geffen Sch. of Med., Univ. of California, Los Angeles, Los Angeles, CA, [16]Dept. of Biostatistics, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Disclosure Block: **M. Zhang:** None.

Identifying disease-critical tissues and cell types has deepened our understanding of disease etiology and may direct efforts to develop treatments (Hakselman et al. 2020 Nat Rev Genet). However, methods that aggregate cells by cell type may fail to capture heterogeneity in scRNA-seq data. Here, we introduce a new approach, single-cell Disease Relevance Score (scDRS), to evaluate the polygenic disease enrichment of individual cells in scRNA-seq data.

scDRS evaluates whether an individual cell has excess expression levels across a set of disease-relevant genes inferred from GWAS, using an appropriately matched empirical null distribution to compute P-values. First, scDRS converts disease GWAS summary statistics into a set of disease-relevant genes using MAGMA (de Leeuw et al. 2015 PLoS Comp Biol). Second, scDRS computes a disease score for each cell as a weighted average expression of disease-relevant genes, with weights inversely proportional to gene-specific technical noise; it also computes a set of control scores for each cell using matched control genes. Finally, scDRS computes disease association P-values for each cell based on the empirical null distribution of control scores across cells. scDRS is well-calibrated in null simulations and well-powered in causal simulations.

We applied scDRS to GWAS summary statistics for 60 diseases and complex traits (average N=297K) in conjunction with Tabula Muris Senis mouse scRNA-seq data (N=300K cells) (The Tabula Muris Consortium 2020 Nature). We reached three main conclusions. First, scDRS results aggregated at the cell-type level recapitulated known biology but also produced novel, biologically plausible findings, associating cardiomyocytes to lymphocyte count and pancreatic beta cells to intelligence. Second, scDRS identified subpopulations of disease-associated cells that are not captured by cell type annotations, including diverse subsets of regulatory T cells and T helper 17 cells associated with inflammatory bowel disease (IBD) and subtypes of neurons associated with schizophrenia (SCZ). Interestingly, scDRS determined that the association with IBD was correlated with effector gradient across T cells (gradient from naÃ¯ve to effector phenotypes) and the association with SCZ was correlated with the spatial dorsal/proximal axis across neurons within the hippocampal CA1 region; scDRS can reveal continuous gradients of cell-disease signals that may be overlooked by cell type-

level analyses. Third, genes whose expression levels across cells were correlated with scDRS scores in disease-critical cell types were strongly enriched for gold-standard drug target and Mendelian disease genes.

PrgmNr 3506 - Resolving the landscape of human genetic variation at the protein domain level using 454,803 exomes from the UK Biobank

View session detail

Author Block: **Y-P. Lai**[1], M. R. Miller[2], C. L. Hyde[3], A. Malarstig[4,5], E. Fauman[2], X. Chen[2]; [1]Early Clinical Dev., Pfizer, Cambridge, MA, [2]Pfizer, Cambridge, MA, [3]Pfizer, Groton, CT, [4]Karolinska Inst., Stockholm, Sweden, [5]Worldwide Res. and Dev., Pfizer, Stockholm, Sweden
Disclosure Block: **Y. Lai:** None.
Recent efforts of sequencing UK Biobank participants have provided us a unique opportunity to understand the human genetic architecture and variation, especially the biological roles of rare, functional variants in a general population. Here, we have systematically summarized the genetic variation in 454,803 humans not only at the transcript level, but also for all protein domains within each transcript. We categorized variation by allele frequencies and functional consequences, including putative loss-of-function (pLoF), missense, in-frame indels, and synonymous variants. We applied a statistical model to estimate cumulative Poisson probability to identify genes with extreme distribution of pLoF variants, and more importantly, to identify domains within each gene that have more or less than expected number of pLoF variants. As a result, we confirmed pLoF-intolerant genes previously reported by gnomAD indeed have lower pLoF mutational diversity rates (defined as the number of functional variants in an exonic region to the length of the region) than the pLoF-tolerant genes. We also evaluated the mutational diversity rates on protein domains and subsequently explored the utility of leveraging this information for identifying optimal units for conducting gene-based association tests using phenotypic traits in the UK Biobank. In sum, we characterized human genetic variation in protein-coding genes and protein domains with a focus on rare, functional variants in exomes. The analysis paves the way for a better understanding of roles of functional variants in protein domains and provides insights to identify and validate new drug targets through human genetics.

PrgmNr 3507 - The Genetic Architecture of Genome-scale Metabolic Networks

View session detail

Author Block: **N. Santhanam**[1], D. Uzuner[2], A. Oommen[3], T. Cakir[2], M. Thambisetty[4], H. Im[1]; [1]The Univ. of Chicago, Chicago, IL, [2]Gebze Technical Univ., Gebze, Turkey, [3]Natl. Univ. of Ireland Galway, Galway, Ireland, [4]NIH, Baltimore, MD
Disclosure Block: **N. Santhanam:** None.
Genome-scale metabolic models seek to reconstruct the totality of biochemical reactions within cells using stoichiometric constraints and mass balance. Integrating genomic, transcriptomic and other omic data into this network promises to provide a more holistic view of the processes that connect them and their role in disease etiology. To date, the genetic architecture underlying individual metabolic reactions in humans is not well understood. Here, we sought to fill this gap by performing genome-wide association study (GWAS), transcriptome-wide association study (TWAS) and heritability studies of genome-scale reaction scores associated with individual metabolic reactions.
We gathered transcriptome data from 255 (GTEx brain) and 336 (Common Mind brain) individuals and reconstructed personalized metabolic models from the latest version of human genome-scale metabolic network, Human-GEM, using the iMAT algorithm. To determine whether 13,416 reactions from GTEx included in Human-GEM were modulated by genetic factors, we then calculated heritability and performed a GWAS analysis of each reaction score. Moreover, we wanted to discern which genes were driving the significant SNP reaction score associations. To do so, we employed PrediXcan analysis, a gene-based association software to compute associations between omic features and a complex trait. With PrediXcan, we predicted expression of GTEx genotype data using prediction models from psychENCODE data that was derived from temporal and frontal cortex data. Predicted expression was then used to determine association with the set of 13,416 metabolic reactions. For the heritability analysis, we used genotype information from the Common Mind Consortium, provided by psychENCODE which included 2,712 reactions. We then determined heritability of these predicted Reaction Scores using a restricted maximum likelihood approach in GCTA.
Our analysis revealed that biochemical reaction scores are genetically modulated. We determined 751 unique SNPs that were associated with reaction scores and 1611 SNP-reaction score pairs at the Bonferroni corrected threshold for p-values. The PrediXcan analysis found a total of nine genes: CYP2D6, CYP2D7, DTX4, FAM109B, LINC00634, NAGA, PKNOX1, SLC25A5P1 and WBP2N that were significantly associated with reaction scores after Bonferroni correction. Moreover, we determined that the reaction scores are indeed heritable with a mean heritability of all reaction scores as 0.037. Understanding these gene-reaction score pairs may further elucidate the genetic regulation of abnormal metabolism underlying complex diseases such as Alzheimerâ□□s disease.

PrgmNr 3508 - The structural landscape of constrained sites in the human proteome

View session detail

Author Block: **B. Li**[1], J. A. Capra[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Univ. of California San Francisco, San Francisco, CA
Disclosure Block: **B. Li:** None.
Quantification of patterns of protein-coding genetic variation within and between species is a cornerstone of evolutionary and functional analyses, and it is a main component of methods for interpreting variants of unknown significance. However, current approaches for quantifying constraint on proteins either focus on individual sites or the whole protein, without accounting for the functional context of the sequence: 3D structure. Recent growth in databases of genetic variation and protein 3D structure enables the synthesis of protein spatial context into the estimation of site-level constraint. Here, we describe a new framework, called COSMIS, for quantification of the constraint on genetic variation in 3D neighborhoods of each protein site based on a mutation-spectrum-aware statistical model of the expected number of variants. We use the COSMIS framework to analyze the distribution patterns of > 1.88 million human standing missense variants across > 2 million sites, covering 47% of all canonical transcripts of the human proteome. We demonstrate that the COSMIS score is accurate in predicting variant pathogenicity and gene essentiality. We further show that COSMIS performs significantly better than a range of 1D sequence-based metrics, such as the MTR score, while also providing biophysical insight into the potential functional roles of constrained sites. To demonstrate the utility of the COSMIS framework, we apply it to detect constrained sites in ion channels and predict the pathogenicity of recently characterized variants using custom-built homology models. We make our constraint maps freely available and anticipate that the structural landscape of constrained sites identified by COSMIS will facilitate interpretation of protein-coding variation in human evolution and prioritization of sites for mechanistic or functional investigation.

PrgmNr 3509 - Transposable element mediated rearrangements are prevalent in human genomes

View session detail

Author Block: **P. Balachandran**[1], I. A. Walawalkar[1], J. I. Flores[1], J. N. Dayton[1], P. A. Audano[1], C. R. Beck[1,2,3]; [1]The Jackson Lab. for Genomic Med., Farmington, CT, [2]Dept. of Genetics and Genome Sci., Univ. of Connecticut Hlth.Ctr., Farmington, CT, [3]Inst. for Systems Genomics, Univ. of Connecticut Hlth.Ctr., Farmington, CT

Disclosure Block: **P. Balachandran:** None.

Transposable elements (TEs) comprise more than half of the human genome, yet their role in genomic instability remains poorly understood. Two TEs can act as substrates for ectopic DNA repair, leading to TE-mediated rearrangements (TEMRs) that differ from retrotransposition events by duplicating or deleting the DNA between TEs. TEMR can result from homologous recombination repair (possibly nonallelic homologous recombination) or non-homologous end joining (NHEJ). Since most TEMR studies have focused on disease-associated events or comparative genomics, genome-wide estimates of TEMR are still not well established. To comprehensively identify TEMRs, we implemented an ensemble pipeline comprising multiple tools and filters on the short-read and long-read whole genome sequencing (WGS) data of three extensively characterized individuals from the Human Genome Structural Variation Consortium. In the three individuals, we identified 4458 nonredundant high-quality deletions and duplications with both long and short read support. Of these variants, 2200 (~50%) were due to retrotransposition and 374 (~8%) were TEMRs, consisting of 300 *Alu-Alu* and 74 L1-L1 rearrangements. Additionally, we found that 30X short-read WGS data is sufficient to identify nearly 90% of deletion and duplication TEMRs. We used a combination of PCR validation and computational techniques to reconstruct breakpoints at nucleotide resolution for the 374 TEMRs, and found that ~90% of deletions and duplications had junctions in TEs on the same strand/orientation of DNA. By mapping the two TEs at the junctions against their corresponding consensus sequences and identifying the position where the left and right breakpoint of a TEMR occurred, we categorized 78% of TEMR as homologous recombination and 22% as NHEJ. This computational approach significantly reduces the burden of manual curation of TEMR, and will be useful in future large-scale studies. We found that *Alu-Alu* and L1-L1 TEMRs occur within regions of relatively high *Alu* and L1 density (p Alu elements are prevalent and are responsible for 80% of TEMRs, they have a notable effect on genes and may contribute to human phenotypic diversity. Through this study, we show that TEs not only affect the genome through retrotransposition but are also a substrate for genetic rearrangements disproportionately affecting genes and may be responsible for phenotypic differences, diseases, and human evolution.

PrgmNr 3510 - Ultra-fast genome-wide association analysis for Alzheimer's Disease in 200,000 exomes from UK Biobank

View session detail

Author Block: **Y-C. Hwang**, C-F. Lin, J. Chin; DNAnexus, Inc., Mountain View, CA
Disclosure Block: **Y. Hwang:** Salary/Employment; DNAnexus.
UK Biobank (UKB), a large prospective population-based initiative, has collected extensive phenotypic data of half a million participants, aged 40-69 years, from the UK. The resource also includes various genetic data, including genome-wide genotype array, whole exome sequencing (WES), and whole genome sequencing (WGS) data for each participant. With this ultra-high dimension data size projected to grow into 15 TB, UKB launches a cloud-based Research Analysis Platform, allowing more than 20K researchers around the world to access them more readily. The platform not only offloads the storage requirement by removing the need of duplicating the data to additional infrastructure, but also provides scalable cloud computing resources. It enables more independent researchers to perform large-scale data analysis such as genome-wide association studies (GWAS), developing advanced machine learning and artificial intelligence techniques. Alzheimerâ□□s Disease (AD) has been causing more than 1 in 9 people aged 65 and older with dementia. There is no effective cure and the care is expensive. In this study, we performed a GWAS on AD using UKB data on the cloud-based platform, starting from data delivered to our UKB application to calculating the summary statistics. From the WES data of 200K samples released lately, we derived a binary trait, AD-by-proxy, based on each participantâ□□s ICD-10 code and their parental diagnoses and ages. We found 22,406 cases and 123,569 controls, after quality controls of the genotype and phenotype data. We applied REGENIE, a machine learning based method, for fitting the whole-genome regression model and testing for associations. Compared with Jansen, et al., we rediscovered variants associated with AD, including the well-known variants in *APOE*, and extended to more coding variants. The entire analysis took less than 12 hours of wall-clock time, which is made possible by both the efficiency of the algorithm and the scalability of the cloud environment. To learn more from the powerful UKB database, we further streamlined this computational-intensive analysis flow as a pipeline, which allows us to extend the analysis to dozens of other phenotypes and traits of interest.

PrgmNr 3511 - Unique Sequence Detection: a novel variant genotyper spanning the full spectrum of genetic variants

View session detail

Author Block: **K. Wong**, A. Rostamianfar, Z. Langley, M. Cusack, A. Ondov, A. Y. Zhou, S. Topper, J. Van Den Akker; Color Hlth., Burlingame, CA
Disclosure Block: **K. Wong:** Salary/Employment; Color Health.
Accurately detecting pathogenic germline variants is crucial for clinical-grade genetic testing. However many remain difficult to detect due to their sequence context or structural complexity. Genotyping has been shown to yield better performance than naive variant calling, particularly with short-read data. The ability to genotype validated and functionally annotated variants can also simplify downstream reporting. Current algorithms, however, are typically limited in their scope of detection. Here, we introduce a novel algorithm, Unique Sequence Detection (USD), to genotype the entire spectrum of variants, ranging from single nucleotide variants (SNVs) to complex structural variants (SVs).

USD operates in two distinct phases: (1) a one-time unique sequence generation step where unique (variant-specific) sequences are constructed based on previously well-characterized variants and (2), a pattern matching step where an alignment file is scanned to identify reads supporting the constructed unique sequences. During the first step, variant-specific sequences are designed algorithmically by evaluating a list of candidate sequences encompassing the variant. Each sequence is constructed of three parts; a â□□coreâ□□ sequence that spans the variant and needs to be matched exactly in the detection step, and left and right â□□flankingâ□□ sequences. The USD algorithm is tolerant of variation in the flanking sequences to provide resilience against co-variants and sequencing errors. For large variants, unique sequences are generated for each breakpoint. If applicable, co-insertions in complex variants are included in the design.

Leveraging clinically reported pathogenic variants, we demonstrate that USD is able to genotype challenging SNVs and indels that were missed or called incorrectly by commonly used variant callers, typically due to high GC content or nearby homopolymers. Over a set of 668 structural variants, USD achieved a recall of >0.9 with a precision >0.99, correctly genotyping deletions, tandem duplications, inversions, insertions, complex rearrangements, and gene conversions. USD especially outperformed other genotypers for deletions/duplications with large co-insertions or resulting from homologous recombination.

Taken together, USD is a comprehensive genotyper that enables accurate detection of variants of all sizes and complexities. It serves to boost detection of known pathogenic variants using either targeted or whole genome short-read sequencing platforms. Most importantly, subsequent observations of validated variants can be added regularly, thereby continuously improving the overall sensitivity of genetic testing.

PrgmNr 3512 - Analysis of concordance of genetic variants across sample types

View session detail

Author Block: **A. Cheng**, Y. Bordiya, L. Manley, T. Jayaweera; Thermo Fisher Scientific, Austin, TX
Disclosure Block: **A. Cheng:** Salary/Employment; Thermo Fisher Scientific.
Recent developments in nucleic acid extraction procedures from diverse sample types have opened the possibility of diagnostic and research applications using various kinds of tissues and body fluids. Currently, several techniques and sample types are being used for nucleic acid extraction, including bronchoalveolar lavage fluid, nasal swabs, brush biopsies, pharyngeal swabs, feces, blood, and saliva. Of these, saliva has emerged as a promising sample type for diagnosis of infectious diseases and genetic applications, including genotyping for personalized medicine, rare disease diagnosis, ancestry prediction, and research on oral cancers. Blood, buccal swabs, and saliva are the three main sample types used for extraction of gDNA to be used for next-generation sequencing (NGS) or qPCR. Blood has historically been the common sample choice for gDNA extraction due to the high quality and quantity of DNA yield. Recently it has become clear that there is an increasing need to utilize noninvasive sample types such as buccal swabs or saliva. Moreover, saliva and buccal swab samples are more economical in their shipping and storage options than other media such as whole blood, serum, and plasma. Saliva samples also require less manipulation compared to blood, serum, plasma, and other sample types. However, data for the concordance between different sample types, especially using the same extraction chemistry and workflows, is limited. To determine if common human genetic variants in a population can be detected consistently across sample types, DNA was extracted from 25 different donors with matched whole blood, buccal swab, and saliva samples. Nucleic acid was extracted and analyzed for quality and quantity of extracted DNA. Concordance among genetic variants across donors was observed using the Ion AmpliSeqâ□¢ Exome panel sequencing workflow and qPCR. We found that extraction from three sample types (whole blood, buccal swabs, and saliva) using the MagMAXâ□¢ DNA Multi-Sample Ultra 2.0 Kit produced high-quality eluates yielding enough DNA for downstream qPCR and NGS applications. When analyzed on the NanoDropâ□¢ Spectrophotometer, eluates showed high purity by consistent A260/A280 and A260/A230 absorbance ratios. Extracted DNA showed consistent genotyping concordance with qPCR, as well as variant calling concordance with whole-exome sequencing across matched sample types. Results from this study show that saliva collection is a cost-effective and noninvasive sampling method that can be used for high-sensitivity downstream applications such as genomic variant calling via qPCR, and NGS with Ion AmpliSeq exome sequencing workflows.

PrgmNr 3513 - Automating Illumina RNA Prep With Enrichment Using Beckman Coulter i7 Hybrid Automated Workstation

View session detail

Author Block: **S. Savant**, R. Pares, M. Giner, R. Nelson; Beckman Coulter Life Sci., Indianapolis, IN
Disclosure Block: **S. Savant:** Salary/Employment; Beckman Coulter.
The Illumina RNA Prep with Enrichment assay is a flexible target enrichment next generation sequencing (NGS) kit for use with targeted RNA sequencing and respiratory viral detection, including detection of the novel SARS-CoV-2 virus. The workflow allows for purified total RNA samples (10-100 ng input) or RNA from formalin-fixed paraffin embedded (FFPE) or degraded samples (20-100 ng input). RNA is reverse transcribed to complementary DNA (cDNA), which is then tagmented using Illuminaâ□□s Enriched Bead-Linked Transposomes to create larger inserts and indexed using IDT for Illumina DNA/RNA Unique Dual (UD) indexes. Libraries can be processed through enrichment by pooling samples down to a 3-plex or as a single-plexed library. The targeted enrichment portion of the assay uses a single 2-hour hybridization to allow for rapid capture and enrichment of the library pools. The entire manual workflow from cDNA synthesis and library preparation to targeted enrichment can be performed in a single day. In this poster, we describe and demonstrate the automated performance of the Illumina RNA Prep, (L) Tagmentation with Enrichment using the Illumina Respiratory Virus Oligos Panel v2 on the Biomek i7 Hybrid Genomics Workstation. The automated method can support cDNA synthesis and library construction between 1 to 96 samples and targeted enrichment and capture between 1 to 96 library pools, allowing users to pool libraries as either a 3-plex or a single-plex. The automated library and enrichment process of 96 samples/32 3-plex pools can prepare libraries for sequencing in approximately 12 hours total time with 45 minutes of hands-on time.

PrgmNr 3514 - Automating Twist Bioscience Modular Library Preparation and Targeted Enrichment Protocols Using Beckman Coulter Automation

View session detail

Author Block: **R. Pares**[1], A. Sahm[2], M. Giner[1], E. Tavkin[1], R. Gantt[2], T. Han[2]; [1]Beckman Coulter Life Sci., Indianapolis, IN, [2]Twist BioSci., San Francisco, CA
Disclosure Block: **R. Pares:** Salary/Employment; Beckman Coulter.
Targeted enrichment of next generation sequencing libraries allows for a cost-effective focus on regions of interest when compared to whole genome sequencing. Hybridization-capture based enrichment is one of the most utilized technologies for targeted enrichment. While hybridization-capture based enrichment can provide a high degree of uniformity and variant detection, it typically involves complex, labor-intensive workflows that can take several days to complete. In this poster, we describe the automation of Twist Bioscienceâ□□s highly modular library preparation and hybridization-capture targeted enrichment chemistries with the use of the Twist Human Core Exome panel on the Beckman Coulter Biomek i7 Hybrid liquid handling platform. The method is capable of preparing up to 96 high quality NGS libraries from human genomic DNA using either enzymatic or mechanical fragmentation workflows and either Twistâ□□s Combinatorial Dual Indices or Universal Adapter System. Libraries can be enriched as either a single plex or pooled together up to an 8-plex using either Twistâ□□s standard or fast hybridization targeted enrichment protocols, resulting in enriched libraries that are ready to be sequenced on Illumina sequencing platforms. All major processes are performed on the system, allowing for maximum user walk away time and a workflow that can be performed in as little as a single day.

PrgmNr 3515 - Automation of NGSâ¯library preparation: Generate high quality libraries with less hands-on time

View session detail

Author Block: **Z. M. Smith**; Beckman Coulter Life Sci., Indianapolis, IN
Disclosure Block: **Z.M. Smith:** Salary/Employment; Beckman Coulter Life Sciences.
Introduction:â¯ Over the past 40 years,â¯massively parallel sequencing has revolutionized DNA sequencing by drastically cutting down the cost of sequencing. As a result, whole genome sequencingâ¯has become a commonplace laboratory practice.â¯However, NGS library preparation is still a labor-heavy process involving multiple manual interactions. Automation of library preparation has the potential toâ¯reduceâ¯hands-on time.â¯ â¯ Methods:â¯â¯ In this poster,â¯we demonstrate the automation of Illumina DNA prep kit onâ¯aâ¯Beckman Coulterâ¯Biomekâ¯NGeniuSâ¯next generation library preparationâ¯system.â¯We loaded 4 human DNA samples into the system with aâ¯10 ngâ¯input concentration. We used an Illuminaâ¯NextSeqâ¯550 sequencer to sequence the libraries andâ¯BaseSpaceâ¯to analyze the results.â¯ â¯ Results: More than 99.2% of reads matched with the reference genome with uniform coverage. The percentages of singletons and duplicates were

PrgmNr 3516 - Germline and somatic protein quantitative trait loci from the tumor proteome in high grade serous ovarian cancer identify candidate biomarkers and connections between germline risk and disease biology

View session detail

Author Block: **K. Dabke**[1], N. Gull[1], P-C. Peng[1], S. G. Coetzee[1], J. T. Plummer[1,2], A. L. P. Reyes[1], B. D. Davis[1,2], S. Chen[1,2], K. Lawrenson[1,3,4], J. Lester[5,6], B. J. Rimel[5,4], A. J. Li[7,4], J-P. B. Govindavari[8], J. K. L. Rutgers[8], B. Y. Karlan[5,6], J. V. Eyk[9,10], S. Gayther[1,2], S. Parker[11], M. R. Jones[1]; [1]Ctr. for Bioinformatics and Functional Genomics, Dept. of BioMed. Sci., Cedars-Sinai Med. Ctr., LOS ANGELES, CA, [2]Applied Genomics, Computation and Translational Core, Cedars Sinai Med. Ctr., Los angeles, CA, [3]Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Inst., Cedars-Sinai Med. Ctr., Los angeles, CA, [4]Div. of Gynecologic Oncology, Dept. of Obstetrics and Gynecology, Cedars-Sinai Med. Ctr., Los angeles, CA, [5]Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Inst., Cedars-Sinai Med. Ctr., LOS ANGELES, CA, [6]Dept. of Obstetrics and Gynecology, Jonsson Comprehensive Cancer Ctr., David Geffen Sch. of Med., UCLA, Los angeles, CA, [7]Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Inst., Cedars-Sinai Med. Ctr., Los Angeles, LOS ANGELES, CA, [8]Dept. of Pathology, Cedars-Sinai Med. Ctr., LOS ANGELES, CA, [9]Dept. of Cardiology, Smidt Heart Inst., Cedars-Sinai Med. Ctr., LOS ANGELES, CA, [10]Advanced Clinical Biosystems Inst., Dept. of BioMed. Sci., Cedars-Sinai Med. Ctr., Los angeles, CA, [11]Advanced Clinical Biosystems Inst., Dept. of BioMed. Sci., Cedars-Sinai Med. Ctr., LOS ANGELES, CA

Disclosure Block: **K. Dabke:** None.

The relationship between germline variation and biology is complex. As our DNA sequence is interpreted through layers of the epigenome, the transcriptome and post transcriptional modification the relationship between variation and these intermediate data layers becomes increasingly complex. While the transcriptome serves as an accessible biomarker that can be easily assayed in tissues, cell types and now single cells, itsâ inconsistent and relatively weak correlation to protein level makes it difficult to ascertain biological insight from changes in the timing or amount of gene expression under disease or treatment conditions. High throughput proteome-wide technologies have improved rapidly in recent years, and mass spectrometry performed using data independent acquisition (MS-DIA) methods now allow for the simultaneous accurate profiling of thousands of proteins from a single sample. This has allowed the relationship between genetic variation and the proteome to be analyzed on genome-wide and proteome-wide scale. Two approaches are available; (i) harnessing a large reference database generated from donors with both proteomics and genetic variation data available for pQTL identification, (ii) applying these newly identified pQTLs to large population scale GWAS cohorts for the identification of genetically predicted protein levels for the identification of proteins associated with disease. In this study we have combined locally generated MS-DIA data from 110 high grade serous ovarian tumors with 169 publicly available tumors from the CPTAC study to identify a set of 398 pQTLS (P

PrgmNr 3517 - Impact of DNA contaminants on next-generation library preparation

View session detail

Author Block: **L. Saunders**, A. Dubinsky; Beckman Coulter Life Sci., Indianapolis, IN
Disclosure Block: **L. Saunders:** Salary/Employment; Beckman Coulter Life Sciences.
Nucleic acid purification and clean-up are critical steps for genomics applications, including library prep and next-generation sequencing (NGS) in genome engineering experiments. Variability in the quality of sample extraction methods can result in differential removal of inhibitors, or reaction contaminants and diminish NGS library preparation quality, as well as downstream data quality. To maximize sample recovery and create more consistent results, efficient removal of inhibitors or contaminants is critical during nucleic acid isolation and purification. In this study, we will use two common DNA extraction technologies (1) magnetic bead-based clean-up, and (2) silica-based columns, for their ability to extract total genomic DNA, and remove contaminants, from human muscle and skin samples. In addition, we will evaluate the effect of residual contaminants such as melanin, collagen and myoglobin on downstream applications such as NGS library prep and read quality metrics. The findings in this study demonstrate the importance of extraction chemistry on sample quality and downstream genomics applications. We also demonstrate the utility of magnetic bead-based- protocols as a superior, reliable, rapid and automatable method to isolate high quality DNA and remove contaminants.

PrgmNr 3518 - Integrated multi-omics analysis of thyroid cancer reveals key molecular pathways involved in tumor formation and metastasis

View session detail

Author Block: **S. Wu**[1], A. Sanghi[1], D. Ramazzotti[1], S. Chen[1], L. A. Orloff[1], M. M. Kasowski[1], M. P. Snyder[2]; [1]Stanford Univ., Palo Alto, CA, [2]Stanford Univ, Stanford, CA
Disclosure Block: **S. Wu:** None.
Thyroid cancer is one of the most common endocrine cancer with a continuously increasing incidence worldwide. Thyroid cancer has the lowest mutational burden among all cancers; however, it exhibits a high degree of heritability. The Cancer Genome Atlas (TCGA) study, one of the most comprehensive studies for thyroid cancer to date, profiled the mutations and transcriptome of 496 primary tumors, and identified two dominant molecular subtypes: BRAF-like and RAS-like. In this study, we performed deep multi-omics profiling (genomics, ATAC-seq, RNA-seq, proteomics, metabolomics, lipidomics) of 27 matched normal thyroid glands, 31 thyroid primary tumors, and 31 lymph node metastasis samples from 36 thyroid cancer patients. We observed dramatic molecular profiling differences between primary tumor/metastasis and normal samples, whereas not between primary tumor and metastasis samples. Transcriptomic, proteomic, metabolomic and lipidomic analyses revealed a dynamic choreography of molecular and cellular events that present three different changing patterns: cancer progression, tumor formation and tumor metastasis. These involve processes such as cell cycle and cell proliferation, cancer-related signaling pathways, immune response, metabolic reprogramming (including Warburg effects, dysregulated energy metabolism, carbohydrates, nucleotides, amino acids and lipids, etc), decreased thyroid hormone biosynthesis. Interestingly, activation of the AhR (aryl hydrocarbon receptor) signaling pathway was found to occur during tumor formation at transcript and protein levels, and we also detected the significant increase of tryptophan catabolite kynurenine, the endogenous tumor-promoting ligand of AhR, from our metabolomics data, indicating AhR signaling pathway and tryptophan catabolism may be the novel drug targets of thyroid cancer. Overall, our results reveal key genomic, molecular and cellular events during thyroid cancer formation and metastasis and its potential molecular mechanisms.

PrgmNr 3519 - Mitigating effects of blood disease pathologies that compromise specificity of gene expression signatures for radiation exposure

View session detail

Author Block: **P. K. Rogan**[1,2], E. J. Mucaki[1], B. Shirley[2]; [1]Western Univ., London, ON, Canada, [2]CytoGnomix Inc, London, ON, Canada
Disclosure Block: **P.K. Rogan:** Major Stockholder/Ownership Interest; CytoGnomix Inc.. Receipt of Intellectual Property Rights/Patent Holder; CytoGnomix Inc..
Background. Combinations of expressed genes can discriminate radiation-exposed from normal control blood samples by machine learning based signatures (with 8 to 20% misclassification rates; PMID 29904591). These signatures can quantify therapeutically-relevant as well as acute accidental radiation exposures. The prodromal symptoms of Acute Radiation Syndrome overlap those present in some viral infections. We recently showed that these radiation signatures produced unexpected false positive misclassification of influenza and dengue infected samples (PMID 33299552). Methods. This study investigated recall by previous and novel radiation signatures independently derived from multiple GEO datasets [GSE6874, GSE10640, GSE85570, GSE102971] on common and rare non-malignant blood disorders and blood-borne infections (thromboembolism [GSE19151], S. aureus infection [GSE30119], malaria [GSE117613], sickle cell disease [GSE35007], polycythemia vera [GSE47018], and aplastic anemia [GSE16334]). Normalized expression levels of signature genes is input to machine learning-based classifiers to predict radiation exposure in other hematological conditions. Results and Discussion. Except for aplastic anemia, these confounders modify the normal baseline expression values, leading to false-positive misclassification of radiation exposures in 8 to 54% of individuals. Shared changes predominantly in DNA damage response and apoptosis-related gene transcripts in radiation and confounding hematological conditions compromise the utility of these signatures for radiation assessment. These confounding conditions are known to induce neutrophil extracellular traps, initiated by chromatin decondensation, fragmentation and often, programmed cell death. Ribovirus infections are proposed to deplete RNA binding proteins, inducing R-loops in chromatin which collide with replication forks resulting in DNA damage and apoptosis. To mitigate the effects of confounders, we evaluated predicted radiation positive samples with novel gene expression signatures derived from radiation-responsive transcripts of secreted blood plasma proteins whose expression levels are unperturbed in these confounding conditions. Conclusions. This two step approach identifies and eliminates misclassified samples with underlying hematological or infectious conditions, leaving only samples with true radiation exposures. Diagnostic accuracy is significantly improved by selecting genes that maximize both sensitivity and specificity in the appropriate tissue using combinations for each of these classes of signatures.

PrgmNr 3520 - Untargeted metabolomics on the diet behaviors and their genomic interaction among Mexican Americans

View session detail

Author Block: **S. Chung**[1], D. Aguilar[1], E. L. Brown[1], C. L. Hanis[1], C. F. Burant[2], C. Evans[2], G. Jun[1]; [1]Dept. of Epidemiology, Human Genetics & Environmental Sci.(EHGES), Sch. of Publ. Hlth., Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [2]Michigan Regional Comprehensive Metabolomics Resource Core, Univ. of Michigan, Ann Arbor, MI
Disclosure Block: **S. Chung:** None.
Diabetes is a gradually progressing disease. 33.9% of adults in the U.S. have prediabetes, but 90% do not know their condition. Prediabetes is reversible by lifestyle changes such as diet. Diet is one of the essential environmental factors contributing to the development of prediabetes, but the effect of nutrient intake on glycemic traits in prediabetes remains unclear. Moreover, the influence of interactions between genes and diet on metabolites needs additional studies. This study aims to determine whether the impact of genes on glycemic traits-related metabolites differ according to diet behaviors among Mexican Americans. A total of 616 individuals were recruited from Starr County, Texas, and 170 samples were analyzed for preliminary analyses. Untargeted plasma metabolomic profiles were analyzed to obtain >6,000 metabolites. All the metabolomic data was standardized and log-transformed. Data for diet behaviors obtained using a semi-quantitative food frequency questionnaire were categorized based on the standards in the fat avoidance scale. Genetic variants were analyzed previously from imputed GWAS data. The association between diet behaviors and metabolites was analyzed by ANOVA model adjusted age, gender, and BMI. The association between metabolites and glycemic traits was also analyzed by linear regression model adjusted age, gender, and BMI. Gene-diet behavior interactions were analyzed using a linear model by PLINK 2.0. Sphingomyelin 32:2 showed a statistically significant association with different kinds of milk consumption (P=3.34E-07). 19 unnamed metabolites showed statistically significant associations with spread, milk, hamburger, and trimfat subgroups;1 metabolite in the spread group, 8 metabolites in the milk consumption group,1 metabolite in the hamburger consumption group, 9 metabolites in trimfat group. Two metabolites associated with trimfat were also associated with glucose intolerance groups and were also highly correlated with leucine and isoleucine (r2>0.7). We found that 36 metabolites associated with fasting glucose, one with 2-hour post glucose load, and 20 metabolites associated with HbA1C. In the Gene-diet behavior interaction analysis, Margaric acid, Eicosatrienoic acid, Cortisol, Lysophosphatidylethanolamine 18:2, and Lauric acids showed the top 5 highest signals (p-value

PrgmNr 3521 - *Best practices for trans-ethnic, meta-analytic transcriptome-wide associations: lessons from the Global Biobank Meta-Initiative*

View session detail

Author Block: **J. HIRBO**[1], A. Bhattacharya[2], W. Zhou[3], E. Gamazon[4], Global Biobank Meta-Initiative Consortium, B. Pasaniuc[5], N. J. Cox[6]; [1]Vanderbilt Univ. Med. Ctr., Clarksville, TN, [2]Univ. of California at Los Angeles, Carrboro, NC, [3]Analytic and Translational Genetics Unit, Massachusetts Gen. Hosp., Boston, MA, [4]VUMC Clare Hall, Univ. of Cambridge, Nashville, TN, [5]Geffen Sch. of Med. at UCLA, Los Angeles, CA, [6]Vanderbilt Univ Med Ctr., Nashville, TN
Disclosure Block: **J. Hirbo:** None.
Large population-based or clinical-case based biobanks offer context to deploy genome-wide associations (GWAS) at scale. The Global Biobank Meta-Initiative (GBMI), through its genetic and demographic diversity, provides a strong opportunity to study population-wide and ancestry-specific genetic associations. A key challenge in GWAS is interpreting and mapping GWAS-significant variants to genes or epigenomic features. Transcriptome-wide association studies (TWAS) are a critical tool to boost detection power for and provide biological context to genetic associations by integrating GWAS signals with predictive genetic models of tissue-specific gene expression. TWAS involves three general steps: (1) training per-gene predictive models of tissue-specific expression in the eQTL dataset using genetic variants, (2) imputing tissue-specific genetically-regulated expression in the GWAS cohort with either individual-level genotypes or summary statistics and a linkage disequilibrium (LD) reference panel, and (3) estimating statistical associations between genetically-regulated expression and trait. Besides what is encountered in GWAS, TWAS presents unique challenges, especially in a trans-ethnic and meta-analytic setting like the GBMI. In this work, we present the GBMI TWAS pipeline in Asthma (n=1,800,785) using 5 tissues in GTEx with genetic and expression data of more than 70 samples from both European and African ancestry samples: subcutaneous adipose (n=71), tibial artery (n=76), skeletal muscle (n=86), sun exposed lower leg skin (n=73), and whole blood (n=80). We explore the performance of the ancestry enriched samples relative to total available samples for these GTEx tissues. We outline a framework for TWAS in a trans-ethnic, meta-analysis across multiple biobanks and explore practical considerations for all the three TWAS steps: ancestry specificity of expression models and LD reference panels, meta-analytic techniques for detection of gene-trait associations, and follow-up tests and analyses for biological context. Our framework can be applied to various phenotypes to study population-wide and ancestry-specific genetic associations mediated by tissue-specific expression. Our work provides a strong foundation for adding tissue-specific gene expression context to biobank-linked genetic association studies, allowing for ancestry-aware discovery to accelerate genomic medicine.

PrgmNr 3522 - Accurate profiling of forensic autosomal STRs using the Oxford Nanopore Technologies MinION device

View session detail

Author Block: **C. Hall**[1], R. Kesharwani[2], N. Phillips[1], F. Sedlazeck[2], R. Zascavage[3]; [1]UNT Hlth.Sci. Ctr., Fort Worth, TX, [2]Baylor Coll. of Med., Houston, TX, [3]Univ. of Texas at Arlington, Arlington, TX
Disclosure Block: **C. Hall:** None.
The high variability characteristic of short tandem repeat (STR) markers is harnessed for human identification in forensic investigations. Despite the power and reliability of current typing techniques, sequence-level information both within and around STRs are masked in the length-based profiles generated. Forensic STR profiling using next generation sequencing (NGS) has therefore gained attention as an alternative to traditional capillary electrophoresis (CE) approaches. In this proof-of-principle study, we evaluate the forensic applicability of the newest and smallest NGS platform available, the Oxford Nanopore Technologies (ONT) MinION device. Although nanopore sequencing on the handheld MinION offers numerous advantages, including on-site sample processing, the relatively high error rate and lack of forensic-specific analysis software has prevented accurate profiling across STR panels in previous studies. Here we present STRspy, a streamlined method able to produce length- and sequence-based alleles designations from noisy, long-read data. To demonstrate the capabilities of STRspy, seven reference samples (female: n = 2; male: n = 5) were amplified at 15 and 30 PCR cycles using the Promega PowerSeq 46GY System and sequenced on the ONT MinION device in triplicate. Basecalled reads were processed with STRspy using a custom STR database containing alleles reported in the STRSeq BioProject NIST 1036 dataset. Resultant STR allele designations and flanking region SNP calls were compared to the manufacturer-validated genotypes for each sample. STRspy generated robust and reliable genotypes across all autosomal STR loci amplified with 30 PCR cycles, achieving 100% concordance based on both length and sequence. Furthermore, we were able to identify flanking region SNPs with >90% accuracy. These results demonstrate that nanopore sequencing platforms are capable of revealing an additional level of variation in and around STR loci with sufficient read coverage. As the first method to successfully profile the entire panel of autosomal STRs amplified by a commercially available kit, STRspy significantly increases the feasibility of using nanopore sequencing in forensic applications.

PrgmNr 3523 - Development of a highly multiplexed RNA in situ hybridization assay for spatial transcriptomic mapping of formalin-fixed paraffin-embedded tissue

View session detail

Author Block: **H. Ji**, M. Yu, H. Lu, L-C. Wang, S. Zhou, B. Zhang, X-J. Ma; Advanced Cell Diagnostics, Newark, CA
Disclosure Block: **H. Ji:** None.
Single cell transcriptomics combined with spatial mapping with RNA in situ hybridization (ISH) holds great promise in resolving heterogeneous tissues at cellular resolution and providing insights into cellular organization and function of diverse cell types in healthy and disease states. ACDâ□□s RNAscopeÂ® represents a major advancement in RNA ISH approaches with its proprietary double-Z probe design and signal amplification that enables highly specific and sensitive target RNA detection. To meet the need for higher multiplexing capability for comprehensive spatial studies dealing with complex and heterogenous tissues, we have previously developed RNAscope HiPlex assay that is capable of multiplex fluorescent detection of up to 12 targets in fresh frozen tissue sections. In this assay, after target probe hybridization and a series of highly effective signal amplifications for all 12 targets, the signals are detected iteratively using a cleavable fluorophore, where in each iteration single RNA transcripts for up to four target genes are visualized as punctate dots in four distinct fluorescent channels.
We present here an upgraded version of the HiPlex assay (HiPlex v2) that can be applied to a wider range of tissue samples including fresh frozen, fixed frozen and formalin-fixed, paraffin-embedded (FFPE) tissues. To achieve this objective, a novel FFPE reagent was developed that can effectively quench tissue background autofluorescence (AF), and thus allowing for a high signal-to-noise ratio to provide single-molecule ISH detection. The novel FFPE reagent minimally affects tissue morphology and signal, and the signal is robust and stable across all iterative rounds for at least up to 2 weeks of assay run. As an example, we demonstrate that HiPlex v2 can detect commonly used immune-oncology markers (such as CD68 and IFNg) in various human cancer tissues, showing its potential for various cancer research applications. Finally, the HiPlex Image Registration Software used to merge images from multiple rounds was also upgraded to incorporate a new background removal algorithm, which provides an additional method of AF removal in registered images. Overall, our new HiPlex v2 assay holds unique advantages over other multiplex fluorescent assays in that up to 12 targets can be detected simultaneously on the same sample, allowing for a streamlined workflow and high performance in complex tissue context. In the new era of single cell biology and spatial transcriptomics, HiPlex v2 assay can be a powerful tool to analyze the spatial organization of cell populations and thus provide a deeper insight into the novel cell populations and their functions.

PrgmNr 3524 - DRIVE: Distant relatedness for identification and variant evaluation

View session detail

Author Block: **J. Baker**[1], H-H. Chen[2], Q. S. Wells[3], D. C. Samuels[4], J. E. Below[5]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Vanderbilt Univ, Nashville, TN, [4]Vanderbilt Univ. Sch. of Med., Nashville, TN, [5]Vanderbilt Univ Med Ctr., Nashville, TN
Disclosure Block: **J. Baker:** None.
Rare variants can be a challenge to accurately detect in many genotyping assays. Identity by descent (IBD), or the concept of pairs of individuals sharing a region of the genome that originated from the same ancestor, can be leveraged to provide additional information that can raise confidence in rare variant genotyping calls in pairs of individuals. In the present study, we used IBD sharing between individuals within Vanderbilt Universityâ□□s biobank, BioVU, to identify networks of people who shared IBD segments around an autosomal dominant pathogenic rare variant. We focused on the rare variant RBM20 Arg636His, which has been shown to cause a Mendelian form of dilated cardiomyopathy (DCM), a disease characterized by cardiac dilation and contractile dysfunction that can lead to heart failure. RBM20 is an RNA binding and splicing regulation protein that is highly expressed in the heart. Participants were identified as carrying the DCM variant on Illuminaâ□□s Mega-Ex array. A network of five individuals was identified as sharing an IBD segment of over 20 centimorgans around the variant. These IBD segments were identified by hap-IBD. All five of these had diagnoses of DCM and other cardiac related phenotypes within their medical records. The presence of the RBM20 variant was confirmed in all five individuals by exome sequencing. A network of two individuals who shared a unique IBD segment within BioVU was also identified. These individuals were genotyped as carrying the DCM variant but only one individual was confirmed by exome sequencing. This methodology allowed us to identify high confidence carriers of the rare variant through IBD sharing around the variantâ□□s genomic location.

PrgmNr 3525 - Enabling faster and more cost-effective drug discovery through flexFS™: A highly durable, elastic, and scalable cloud-native file system

View session detail

Author Block: **G. Planthaber**[1], M. Colosimo[1], A. Poliakov[1], M. Matz[1], S. Sarangi[2], Z. Pitluk[1]; [1]Paradigm4, Waltham, MA, [2]Waltham, MA
Disclosure Block: **G. Planthaber:** Salary/Employment; Paradigm4. Receipt of Intellectual Property Rights/Patent Holder; Paradigm4.
The proliferation and availability of multi-omics and single cell data along with cutting-edge analytical tools is dramatically changing the drug and biomarker discovery landscape. As datasets become larger and algorithms become more data and computationally intensive, there is a complementary need for faster, more flexible, and less expensive data storage. Often multi-disciplinary teams need concurrent access to these data in a reliable, efficient, secure and durable fashion. Data must be encrypted and allow access control while supporting analytics and machine learning clusters that need rapid streaming of data. In this poster we introduce flexFS™, an elastic cloud-native file system that offers the cost effectiveness of S3 with the ease-of-use of a POSIX file system. flexFS seamlessly services 1000â□□s of hosts across the network. As an illustration of scalability, we present results from a large burden test performed with UK Biobank data and run using the REVEAL™: Biobank application backed by flexFS. The burden test was performed on 52,821 genes comprising of 5,091,192 variants, 541 phenotypes, and used 14 covariates. flexFS also provides accessible and reproducible archival and reference data storage. In addition to discussing performance details, we will present time and cost comparisons and present additional scenarios for utilization of flexFS.

PrgmNr 3526 - Enabling Real Time Analysis of Single-Cell Genomic Data with Accelerated Computing

View session detail

Author Block: **A. Lal**, C. Nolet, R. Ilango, T. Dyer, J. Zedlewski, J. Israeli; NVIDIA, Santa Clara, CA
Disclosure Block: **A. Lal:** None.

Single-cell sequencing experiments measure the molecular properties of individual cells, enabling high-resolution genomic and transcriptomic studies. Analysis of single-cell data requires high speed and interactivity to allow easy data exploration, clustering and visualization at different scales, and detailed comparisons between cell types. The growth of data has made analysis increasingly difficult, with datasets of over 1 million cells now possible.

Here we present open-source Python pipelines that dramatically accelerate single-cell data analysis. Our pipelines are built upon RAPIDS (https://rapids.ai), a free, open-source software suite for GPU-accelerated data science.

As a demonstration, we analyzed single-cell RNA-seq data from 1 million mouse brain cells. Compared to a state-of-the-art pipeline using the Scanpy library on 32 CPU cores, we achieved significant acceleration for every step, including 178x faster t-SNE visualization, 98x faster UMAP, and 2000x faster Leiden clustering, on a Tesla A100 GPU. We were able to cluster all cells in under 2 seconds and run the complete analysis in under 10 minutes, compared to 3.5 hours. We also developed a GPU-accelerated single-cell ATAC-seq pipeline that includes many of the same steps.

To make data exploration even easier, we developed an open-source GPU-powered cell browser that runs within a Jupyter notebook. In this browser, users can apply RAPIDS to visualize a single-cell dataset, select groups of cells, and analyze the selected cells. To our knowledge, this is the first tool to enable real-time, in-browser single-cell analysis at scale. Finally, we enabled AtacWorks, a deep learning tool to enhance epigenomic data, to run interactively as part of our single-cell ATAC-seq pipeline. This demonstrates how deep learning can be integrated with RAPIDS to provide end-to-end GPU pipelines that improve accuracy as well as speed.

Our open-source tools allow researchers to explore massive datasets in real time, and to iterate rapidly over algorithms and parameters to achieve better results. All of our demonstrations are publicly available as Jupyter notebooks at https://github.com/clara-parabricks/rapids-single-cell-examples.

PrgmNr 3527 - Improving rare variant calls from genotype array data: the Million Veteran Program Return Of Actionable Results (MVP-ROAR) Study

View session detail

Author Block: **Y. Shi**[1], M. Danowski[1], P. Devineni[1], D. Dochtermann[1], J. L. Vassy[2], S. Pyarajan[3], Million Veteran Program; [1]VA Boston Hlth.care System, Boston, MA, [2]Harvard Med. Sch. at VA Boston Hlth.care System, Boston, MA, [3]

Disclosure Block: **Y. Shi:** None.

Background: Familial hypercholesterolemia (FH) is an autosomal dominant condition characterized by low-density lipoprotein cholesterol (LDL-C) elevation and early cardiovascular disease. Despite an estimated prevalence of 1:311, ~90% of individuals with FH remain undiagnosed. Ongoing curation efforts are improving the ability to identify pathogenic variants associated with FH, most commonly in the *LDLR* gene. For this reason, the Million Veteran Program (MVP) launched the Return Of Actionable Results (ROAR) trial to study the processes and outcomes of identifying, verifying, and reporting FH-associated variants to MVP participants from their MVP chip data (Axiom Biobank Array of ~723K variants enriched for known disease-associated SNPs). However, among the first 8 participants with FH-associated variants identified from genotyping, only 5 participantsâ variants were confirmed by sequencing at a commercial laboratory. Here we describe the development and implementation of a novel quality control (QC) process to improve the detection of FH-associated variants in MVP genotype array data. **Methods**: We considered a list of 36 *LDLR* variants classified as pathogenic or likely pathogenic by a commercial laboratory, using American College of Medical Genetics and Genomics-Association for Molecular Pathology (ACMG-AMP) criteria. We developed a QC process to validate rare heterozygous genotypes (rareHets) called on the MVP chip. QC parameters were generated by the Rare Het Adjustment (RHA) Algorithm, which was triggered when fewer than four rareHets were called per batch of ~4500 participants. Next-generation sequencing (NGS) research data were available for 7 participants; a support vector machine (SVM) was built to minimize false positive calls, using genotypes from NGS as truth. Results were additionally compared to those from ROAR participants who had undergone commercial laboratory confirmation by gene sequencing. **Results**: Out of 529 *LDLR* rareHets initially detected by standard genotype calling among 469,170 MVP participants, the RHA Algorithm removed 151 as NoCall. Among the remaining 378, the SVM predicted 300 as high confidence calls. 100% call concordance between the QC pipeline and NGS was observed in all 7 participants with available research NGS data and between the QC pipeline and commercial gene sequencing for all 4 participants with confirmatory test results. **Conclusions**: We developed a QC process to reduce the rate of false positive rareHet calls from genotype array data. This improved analytic validity is a critical step before considering returning results to participants and to realize the potential of population precision health.

PrgmNr 3528 - Integration of WGS & RNA-seq as an efficient approach for novel variant discovery in complex genetic cases

View session detail

Author Block: **S. Audet**[1,2], E. Bareke[2], C. Michaud[1,2], V. Triassi[1,2], N. Legault-Cadieux[1], L. Touma[3], V. Ferraro[3], A. Duquette[2,3], M. TÃ©treault[1,2]; [1]Univ. of Montreal, Montreal, QC, Canada, [2]CHUM Res. Ctr., Montreal, QC, Canada, [3]CHUM, Montreal, QC, Canada
Disclosure Block: **S. Audet:** None.
This pilot study aims to develop an integrative analysis method that allows for an increased diagnosis success rate of rare genetic mutations. Moreover, identification of novel genes associated with Episodic Ataxia (EA) and evaluation of new AI-generated prediction algorithm, for a more robust variant examination, will ensue from the investigation.
Characterized by sporadic loss of voluntary movement coordination, EA typically manifest with a late onset as well as high clinical and genetic heterogeneity, setting additional hurdles to diagnosis. While four genes have been linked to the eight subtypes of EA, many patients are left without molecular diagnosis due to the limitations of individual DNA-sequencing methods, which can be mitigated by the functional overview that RNA-seq offers.
EA patients lacking molecular diagnosis despite in-depth examination were recruited in Montreal. Whole-Genome and RNA-sequencing was performed on blood samples to identify variants, differential expression, splicing events and repeat expansion. Multiple recent pathogenicity prediction algorithms were chosen to be tested concurrently to standard ones.
Candidate variants were identified for each patients according to pathogenicity scores, rarity of the genetic events, and known functional as well as clinical information for a given altered gene. Among the findings are truncations, missenses, and alternative splicing in genes already associated to either Spinocerebellar Ataxia or Spastic Paraplegia. In addition to being present in both datasets, validation of these interesting genomic events has been performed through Sanger Sequencing of both DNA and RNA. The available functional information from RNA-seq suggests abnormal mRNA expression, and is supported by the Sanger Sequencing as well as a traditional qPCR. A meta-analysis of our patientsâ transcriptomic profiles could also uncover commonly affected pathways in EA development.
This project should provide more defined diagnosis, leading to better quality of life, better understanding of prognosis and better management of care for patients. It will also promote the integrative approach for a larger spectrum of disorders and might eventually lead to new therapeutic strategies.

PrgmNr 3529 - Introduction of a walk-away automated Roche NGS workflow solution: Integrated KAPA library preparation, KAPA target enrichment and the AVENIO Edge instrument

View session detail

Author Block: **P. Wadia**, E. Malfarta, A. Liang, J. Lefkowitz, S. Sharma, D. Emani, P. Ton, B. Morck, T. Williams, J. Dasgupta, A. Fararooni, A. Moghaddasi, A. Al-Ariemy, M. Pintor, B. LaRochelle, N. Razavi, P-L. Janvier; Roche, Pleasanton, CA
Disclosure Block: **P. Wadia:** Salary/Employment; Roche.
Introduction: By automating and simplifying NGS workflow steps, a variety of diagnostic applications become practical and robust, thus increasing the efficiency of precision medicine. The AVENIO Edge* instrument provides a complete walk-away automated NGS library prep solution with on-deck QC from extracted nucleic acid with minimal hands-on time and flexibility for KAPA Target Enrichment. Feasibility of an NGS workflow with zero pipetting steps using Roche KAPA Library Preparation and KAPA Target Enrichment reagents with an AVENIO Edge instrument was explored. **Methods:** Performance, carry-over contamination rate, turn-around time, DNA quantitation module testing, were representative experiments comparing automated vs. manual capture workflows using Roche KAPA HyperCap Workflow v3.0 reagents on the AVENIO Edge instrument. Representative panels including KAPA HyperExome Panels were tested with pre-capture and post-capture pooling workflows using genomic DNA. Sequencing data was analyzed through internal pipelines and included percent reads on-target, fold-80 base penalty, mean target coverage, total duplicate rate, 90th/10th percentile ratio, mean insert size. **Results:** The AVENIO Edge Quant kit results and sequencing metrics were similar between automated and manually prepared samples for all representative panels. No significant differences were observed from reagent lot-to-lot, run to run & between instruments with minimal contamination across runs. Average turn-around-time for 24 sample processing was ~31 hours. **Conclusions:** We successfully demonstrated the integration and performance of the complete Roche KAPA HyperCap Workflow V3.0 leveraging KAPA Library Preparation and KAPA Target Enrichment reagents on the AVENIO Edge instrument, enabling the broader adoption of NGS in precision medicine and ultimately improving patient outcomes. *AVENIO Edge is currently in development. The data presented here are not intended for diagnosis or patient management. For Research Use Only. Not for use in diagnostic procedures.

PrgmNr 3530 - Investigation of a dysmorphic facial phenotype in patients with Gaucher type 2 and 3

View session detail

Author Block: **E. Daykin**[1], N. Fleischer[2], M. Abdelwahab[3], N. Hassib[4], R. Schiffmann[5], E. Ryan[1], G. J. Lopez[1], E. Sidransky[1]; [1]Med. Genetics Branch, NHGRI, NIH, Bethesda, MD, [2]FDNA Inc., Boston, MA, [3]Cairo Univ. Pediatric Hosp., and Social and Preventive Med. Ctr., Kasr Elainy Hosp., Cairo, Egypt, [4]Orodental Genetics Dept., Human Genetics and Genome Res. Div., Natl. Res. Ctr., Cairo, Egypt, [5]Baylor Res Inst, Dallas, TX

Disclosure Block: **E. Daykin:** None.

Gaucher disease (GD) is an autosomal recessive lysosomal storage disorder usually classified into three distinct types: non-neuronopathic (GD1), acute neuronopathic (GD2) and chronic neuronopathic (GD3). Discriminating between the types in a newly diagnosed child can be challenging due to the vast heterogeneity of clinical presentations and the limited genotype-phenotype correlation. Patients with GD2 often face a long diagnostic odyssey given the rarity of this disease. New technologies have enabled pilot newborn screening programs and widespread parental carrier testing, facilitating earlier diagnoses but exacerbating the need for prompt discrimination between the Gaucher types in order to implement appropriate management and establish prognosis. Interestingly, parents of babies with GD2 and physicians treating patients with GD3 have observed that affected children often share facial features. DeepGestalt is a deep learning algorithm that analyzes patient photos using facial recognition algorithms employing artificial intelligence to generate lists of genetic syndromes with similar morphologies. Analyzing 103 photos of 47 patients with GD2 and 143 photos of 86 patients with GD3, we generated receiver operating characteristic (ROC) curves comparing patient photos to age, sex, and ethnicity-matched controls. We then compared our GD2 cohort to our GD3 cohort through binary analysis and ran a multiclass comparison of our GD2 and GD3 cohorts, along with 133 photos of 48 patients with GD1. All areas under the curve (AUC) of the ROC curves were $\geq$ 0.89 (p values

PrgmNr 3531 - Multiple *de novo* copy number variant (M*dn*CNV) driven mirror traits and blended phenotype

View session detail

Author Block: **H. Du**[1], A. Jolly[1,2], C. M. Grochowski[3], B. Yuan[4], M. Dawood[5], S. N. Jhangiani[3], H. Li[3], D. M. Muzny[3], J. M. Fatih[5], Z. Coban-Akdemir[6], M. E. Carlin[7], A. E. Scheuerle[8], J. E. Posey[5], M. Pendleton[9], S. Juul[9], P. J. Hastings[1], W. Bi[1,10], R. A. Gibbs[3], J. R. Lupski[11], F. J. Sedlazeck[3], C. M. Carvalho[12], P. Liu[13,10]; [1]Molecular and Human Genetics, Baylor Coll. Med., Houston, TX, [2]Med. Scientist Training Program, Baylor Coll. of Med., Houston, TX, [3]Baylor Coll. Med., Houston, TX, [4]Seattle Children's Hosp., Seattle, WA, [5]Baylor Coll. of Med., Houston, TX, [6]Human Genetics Ctr., Dept. of Epidemiology, Human Genetics, and Environmental Sci., Sch. of Publ. Hlth., The Univ. of Texas Hlth.Sci. Ctr. at Houston,, Houston, TX, [7]Gen & Dev Ctr Southwest, Arlington, TX, [8]Univ. of Texas, Southwestern Med. Ctr., Dallas, TX, [9]Oxford Nanopore Technologies, New York, NY, [10]Baylor Genetics Lab., Houston, TX, [11]Baylor Col Med, Houston, TX, [12]Pacific Northwest Res. Inst., Seattle, WA, [13]Molecular and Human Genetics, Baylor Coll. of Med., Houston, TX

Disclosure Block: **H. Du:** None.

A unique mutational phenomenon described by Liu *et al*. highlighted individuals with variable congenital abnormalities and multiple (n â¥ 4), large (>100 kb), *de novo* copy number variants (M*dn*CNV) throughout their genomes. The M*dn*CNV phenotype is ultra-rare and detected in 5/60,000 individuals referred for genome-wide clinical microarray. The variable congenital abnormalities observed in individuals with a M*dn*CNV phenotype are thought to be caused by the copy number change of different critical driver gene(s). To understand the mutational mechanism underlying the M*dn*CNV phenomenon, we leveraged long (LR-WGS) and short (SR-WGS) read whole genome data with integration and visualization tools to characterize the mutational signatures of *de novo* mutations (DNM) using a trio approach in an unpublished M*dn*CNV family. Eight *dn*CNVs of average length ~1 Mb were mapped in the proband. Biparental origin of constitutive *dn*CNVs (4:4) supports early perizygotic mutagenesis as the cause of genome-wide hypermutation. Sequence microhomology/microhomeology was present at 6/8 breakpoint junctions. *De novo* SNVs (6/79) and *de novo* indels (1/12) are enriched within 4 Mb of the *dn*CNV regions, suggesting microhomology-mediated break-induced replication, a replicative recombination-repair, as the process underlying M*dn*CNV formation. Gene content of the affected genomic regions were analyzed using the human phenotype ontology (HPO) to identify potential driver gene(s). Of the genes encompassed by *dn*CNVs, our proband has highest phenotype similarity score with published reports of *NSD1* and *SMARCC2* variants, suggesting the observed trait manifestation may be driven by gene dosage effects at multiple loci as a form of multilocus pathogenic variation (MPV). Of the known phenotypic traits of *NSD1* variants, our proband has opposite features, i.e. mirror traits, compared to individuals with a diagnosis of *NSD1* deletion syndrome. Our study highlights the multimodal genomic analysis approach in characterizing DNM. Moreover, we demonstrate the utility of quantitative phenotypic analysis to identify contributory, disease-associated genes within a background of genome-wide *dn*CNV, and provide evidence for duplications at two genomic loci containing triplosensitive genes that contribute to the probandâ□□s blended phenotype.

PrgmNr 3532 - Rapid endpoint PCR: Improving amplification speed without sacrificing specificity

View session detail

Author Block: **A. Waite**, A. Taft; Promega, Fitchburg, WI
Disclosure Block: **A. Waite:** Salary/Employment; Promega.
End-point PCR is essential for gene amplification and its global use continues to grow alongside other popular PCR technologies (qPCR, ddPCR, etc.). Within large-scale screening studies, speed to result is becoming increasingly important. Recent advancements in instrumentation have led to potential improvements in turnaround time, but limitations in reagent capabilities and speed remain a roadblock.
This study describes the development of a new reagent formulation targeted towards faster end-point PCR. Currently, the conventional PCR process may take 90 minutes. Utilizing the new formulation with â□□fastâ□□ thermal cycler programming can improve timing to 30 minutes or less. We will demonstrate a newly optimized master mix that is able to reduce the reaction time to

PrgmNr 3533 - Sequence-based assessment of mitochondrial DNA oxidative damage in cognitive impairment: Shedding light on health disparities in Mexican Americans

View session detail

Author Block: **D. M. Reid**[1], R. C. Barber[2], Z. Zhou[3], N. R. Phillips[1]; [1]Microbiol., Immunology, and Genetics, Graduate Sch. of BioMed. Sci., UNT Hlth.Sci. Ctr., Fort Worth, TX, [2]Graduate Sch. of BioMed. Sci., UNT Hlth.Sci. Ctr., Fort Worth, TX, [3]Biostatistics & Epidemiology, Sch. of Publ. Hlth., UNT Hlth.Sci. Ctr., Fort Worth, TX

Disclosure Block: **D.M. Reid:** None.

The aging population (65+) is rapidly expanding, leading to increased prevalence of age-related diseases (e.g., cardiovascular disease, metabolic disorders, cancer, and neurodegenerative diseases). Alzheimerâ□□s Disease (AD) is the 3rdleading cause of death in aging adults in the U.S., and age is the greatest risk factor. The U.S. Hispanic/Latinx population is expected to substantially expand through the year 2060 compared to their non-Hispanic/Latinx white (NHW) counterparts, resulting in the projected number of AD cases in the Hispanic/Latinx population to quadruple. Distinctively in the Mexican American (MA) population, diabetes, depression, stroke, and obesity are prevalent risk factors for developing cognitive impairment; however, reasons for the association between cognitive decline and these comorbidities remain unclear. Studies have shown correlations between commonly observed AD pathological changes and DNA damage. The mitochondrial genome is particularly vulnerable to DNA damage due to its close proximity to reactive oxygen species (ROS), which accumulate with age. Lifestyle and/or metabolic health may contribute directly to age-related risk for neurodegeneration. We are interested in determining if abnormal mitochondrial function, as indicated by oxidative DNA damage, is associated with cognitive impairment in MAs. Oxidative damage to guanine (G) forming 8oxoG, is one of the most prevalent DNA lesions. Here, we investigate the mtDNA 8oxoG mutational load in participants of the Texas Alzheimerâ□□s Research and Care Consortium (n=577) using Illumina-based NGS. We compare the 8oxoG load among in MAs and NHWs with AD, mild cognitive impairment, type-2 diabetes (T2D), and/or comorbidity (cognitive impairment with T2D). Our data indicates 8oxoG mutational load is significantly higher in MAs than in NHWs and is associated with cognitive function, sex, and education. Analysis of 8oxoG mutational â□□hotspotsâ□□ were not significantly associated with cognitive phenotypes in MAs or NHWs. Stratified analysis for 8oxoG mutational load in MAs suggests significant elevation when comparing cognitive impairment to normal controls. For future work, we propose nanopore sequencing technology as an improved alternative to current detection and quantification methods and provide preliminary proof of concept results for this method in analysis of oxidative mtDNA damage. Understanding the extent and implications of oxidative mtDNA damage may aid our understanding of the differences in manifestation of age-related cognitive decline in MAs as compared to NHWs.

PrgmNr 3534 - TowardsIsoform Resolution Single-Cell Transcriptomicsfor Clinical Applications Using Highly Accurate Long-Read Sequencing

View session detail

Author Block: **E. Tseng**[1], J. G. Underwood[1], A. S. Nanda[2], V. Ramani[2], S. Furlan[3]; [1]Pacific BioSci.s, Menlo Park, CA, [2]UCSF, San Francisco, CA, [3]Univ. of Washington, Seattle, WA
Disclosure Block: **E. Tseng:** Salary/Employment; Pacific Biosciences.
Single-Cell RNA-Seq (scRNA-Seq) emerged for the characterization of gene expression differences between different individual cells, which holds promise in clinical applications such as cancer, neurodegenerative, and immunological diseases. However, to date most scRNA-Seq methods employ short-read sequencing platforms that can only offer gene-level information, which would miss critical isoform information that could be markers for disease or development.
PacBio long-read transcript sequencing, (Iso-Seq method) can characterize full-length isoforms and has been used in bulk RNA studies to identify alternative splicing, fusion genes, and allele-specific isoform expression. Recently, the Iso-Seq method has been applied to the 10X Genomics single-cell platform to study postnatal mouse brain regions. This work identified distinct splicing patterns that differentiated by cell type, a phenomenon that would have gone undetected using short reads.
To explore the Single-Cell Iso-Seq method for potential enablement of clinical applications in a cost-efficient manner, we evaluated two techniques to improve the yield of single-cell molecules on the 10X Genomics platform. We tested our method on two human PBMC samples and showed that we can obtain a 6-fold throughput increase of 8-9 million full-length cDNA molecules per SMRT Cell 8M on the Sequel IIe System with identifiable barcodes and unique molecular identifiers (UMIs). The long-read data showed near perfect concordance with matching short read data on captured single cells and molecule abundances. We were also able to recapitulate similar cell type clustering as the short reads. The full-length isoform information revealed distinct expression levels in T cells and a wide diversity of splicing patterns not observable through 3â□□ tagging methods. Furthermore, as with bulk Iso-Seq analysis, allele-specific isoform expression and full-length transcript-based ORF prediction are possible and could lead to important discoveries for clinical applications.

PrgmNr 3535 - Attitudes on pharmacogenetic results as secondary findings among medical genetics providers

View session detail

Author Block: **M. N. Bartos**[1,2], S. A. Scott[1,3], E. W. Jabs[1], H. Naik[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]The Univ. of Alabama at Birmingham, Birmingham, AL, [3]Stanford Univ., Palo Alto, CA
Disclosure Block: **M.N. Bartos:** None.
As evidence increases supporting the potential utility of pharmacogenomics and the growing interest in implementing pharmacogenomic-guided prescribing, incorporating pharmacogenomic variants as secondary findings into clinical sequencing tests is increasingly under consideration. Therefore, we explored clinical provider perceptions on the utility of receiving pharmacogenomic results as secondary findings, specifically evaluating internal medicine and pediatric-geneticist attitudes towards it and their comfort with comprehension, interpretation, and result translation. Four focus groups were conducted with 12 geneticists, and standard thematic analysis conducted. All participants had experience ordering sequencing tests; however, the majority (10/12) did not or rarely order pharmacogenomic tests or prescribe medications with established response variability. Half of the participants had low comfort interpreting results without the support of clinical resources, whereas the other half were comfortable interpreting results and navigating available resources independently. The most frequently discussed positives of receiving pharmacogenomic results as secondary findings included prevention of adverse drug reactions in adults; grateful patients who are information seekers; the ability to prescribe more effective treatments rapidly; and appreciation of great advances in knowledge and guidelines of pharmacogenomics in recent years. Negatives included laboratory reporting issues, such as not following guidelines or an incorrect interpretation provided on a report; laboratory marketing to individuals without prior pharmacogenetic knowledge or targeting populations; pharmacogenomics findings buried in lengthy reports in patientsâ charts; and pharmacogenomics testing only applicable to select populations such as adults or patients treated with certain medications. The most desirable pharmacogenomic resources discussed were the creation of an electronic health record clinical decision support tool to assist with identifying and implementing pharmacogenomic results; a specialized pharmacist as part of the care team; more pharmacogenomic training during medical/graduate school; and well-written interpretation of pharmacogenomic results included on laboratory reports. Importantly, this novel focus group study determined that a majority of participants (10/12) agreed that it is reasonable to consider adding certain actionable pharmacogenomic genes to the ACMG secondary findings list; however, this was qualified with a need for additional resources supporting implementation.

PrgmNr 3536 - Connectivity Map Analysis of a Single-Cell RNA-Sequencing Derived Transcriptional Signature of mTOR Signaling

View session detail

Author Block: **N. Mahi**, M. Medvedovic, J. Yu; Univ. of Cincinnati, Cincinnati, OH
Disclosure Block: **N. Mahi:** None.

In the connectivity map (CMap) approach to drug repositioning and development, transcriptional signature of disease is constructed by differential gene expression analysis between the diseased tissue or cells and the control. The negative correlation between the transcriptional disease signature and the transcriptional signature of the drug, or a bioactive compound, is assumed to indicate its ability to â□□reverseâ□□ the disease process. A major limitation of traditional CMap analysis is the use of signatures derived from bulk disease tissues. Since the key driver pathways are most likely dysregulated in only a subset of cells, the â□□averagedâ□□ transcriptional signatures resulting from bulk analysis lack the resolution to effectively identify effective therapeutic agents. The use of single-cell RNA-seq (scRNA-seq) transcriptomic assay facilitates construction of disease signatures that are specific to individual cell types, but methods for using scRNA-seq data in the context of CMap analysis are lacking. Lymphangioleiomyomatosis (LAM) mutations in TSC1 or TSC2 genes result in the activation of the mTOR complex 1 (mTORC1). The mTORC1 inhibitor Sirolimus is the only FDA-approved drug to treat LAM. Novel therapies for LAM are urgently needed as the disease recurs with discontinuation of the treatment and some patients are insensitive to the drug. We developed methods for constructing disease transcriptional signatures and CMap analysis using scRNA-seq profiling and applied them in the analysis of scRNA-seq data of lung tissue from naÃ¯ve and sirolimus-treated LAM patients. New methods successfully implicated mTORC1 inhibitors, including Sirolimus, as capable of reverting the LAM transcriptional signatures. The CMap analysis mimicking standard bulk-tissue approach failed to detect any connection between the LAM signature and mTORC1 signaling. This indicates that the precise signature derived from scRNA-seq data using our methods is the crucial difference between the success and the failure to identify effective therapeutic treatments in CMap analysis.

PrgmNr 3537 - Explainable machine learning for adverse drug reactions

View session detail

Author Block: **C-F. Lin**, P. Ngyun, D. Quang, J. Chin; Res. Lab, DNAnexus Inc., Mountain View, CA
Disclosure Block: **C. Lin:** Salary/Employment; DNAnexus Inc..
The advancement of pharmacogenomics (PGx) has led to the establishment of gene-drug specific dosing guidelines. Implementing such guidelines in clinical settings can reduce the incidence of adverse drug reactions (ADRs). Unfortunately, dosing guidelines are only available for a small fraction of the over 20,000 prescription drug products approved for marketing. Furthermore, while some ADRs are predictable from the drugâ□□s known pharmacology, others are not. UK Biobank, with its wealth of real-world clinical data, questionnaires, and genomic data for nearly half-million volunteers, provides a valuable resource for studying ADRs. We set out to do this by building and explaining machine learning models for predicting ADR risk for individuals.
We included anthropometric measurements, verbal interview questions related to medical history, lab test results, and ICD10 codes in the main and secondary diagnoses. We devised an algorithm to select ICD10 codes and transformed them into features for modeling. For genomic features, we focused on rare variants within characterized protein domains in the known â□□ADMEâ□□ (absorption, distribution, metabolism, and excretion) genes and â□□Very Important Pharmacogenesâ□□ curated by PharmGKB. We also included Polygenic Risk Scores calculated using a number of conditions and data we selected from the PGS catalog.
We chose LightGBM (a decision-tree based gradient boosting framework) as our tool for modeling. We applied the SHAP (SHapley Additive exPlanation) package to the modeling results. SHAP explains the learned model by estimating how much does one feature impact the prediction results. Not only does this indicate which features are potentially more relevant, it also reveals unintended structure in the data, which we encountered and had to rectify in the process. Among the top impactful features, there were assorted blood assay results and number of medications, etc. The top two, however, were related to mental states.
Additionally, we have explored the use of admixture models on rare variants to visualize and identify sets of genes that are involved in specific ADR types in a concerted way. In particular, we used CountClust, which was developed for RNA-seq analyses and uses Grade of Membership models (aka Latent Dirichlet Allocation).

PrgmNr 3538 - Identifying the genetic contributors of efficacy and adverse metabolic effects of thiazide diuretics in African Americans from the Genetics of Hypertension Associated Treatments (GenHAT) study

View session detail

Author Block: **N. M. Davis Armstrong**[1], V. Srinivasasainagendra[1], R. M. Tanner[1], B. A. Hidalgo[1], H. K. Tiwari[1], D. K. Arnett[2], M. R. Irvin[1]; [1]Univ. of Alabama at Birmingham, Birmingham, AL, [2]Univ. of Kentucky, Lexington, KY
Disclosure Block: **N.M. Davis Armstrong:** None.
African Americans (AAs) have a higher prevalence of hypertension (HTN) compared to whites, presenting with a more severe form due to earlier onset and more rapid vascular damage. AAs respond better to diuretics compared to beta blockers or ACE inhibitors, yet the reason is not well understood. While thiazide diuretics continue to be a first-line antihypertensive, there is clinical significance to adverse metabolic effects linked to thiazide use, including incident diabetes and hypokalemia. The GenHAT study is an ancillary study of the Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). DNA from participants was extracted and hybridized to Illumina Multi-Ethnic AMR/AFR BeadChip arrays. Quality control was performed at the sample and variant level, resulting in the inclusion of 4,297 AAs taking chlorthalidone, a thiazide diuretic, and 969,031 genotyped variants. Upon imputation using the NHLBI TOPMed Freeze 8 reference panel, over 20 million variants with minor allele counts >20, imputation quality scores >0.3, and genotype probabilities >90% were retained for association analysis. Outcomes of interest included systolic (SBP) and diastolic (DBP) blood pressure response over 6 months, fasting glucose response (FG) over 24 months, and serum potassium response (K) over 2 months. Linear regression models for the response of all outcomes were performed in PLINK2 and adjusted for age, sex, baseline measure, and genetic ancestry. A total of 14 variants (3 SBP, 2 DBP, and 9 FG) exceeded statistical significance at pCDHR2 and *MINDY3-CUBN* gene regions for SBP, *LINC02211-CDH9* intergenic region for DBP, and *GIMAP1-GIMAP5* for FG analyses. While many of the variants identified had no previous association with HTN or drug response, variants located within *CDH9*, *MINDY3-CUBN*, and *GIMAP* gene regions were of particular interest. *CDH9* variants have been previously associated with coronary artery calcification in an AA meta-analysis, while variants in *CUBN* have been linked to albuminuria and coronary artery disease. The *GIMAP* GTPase family are potential modifiers in autoimmune diseases including diabetes, and have reported associations with triglycerides, von Willebrand factor, C-reactive protein, and fibrinogen levels. Replication for these variants is ongoing using data from the International Consortium for Antihypertensive Pharmacogenomics Studies (ICAPS) to determine whether these variants can help identify individuals at risk of poorer drug response or unfavorable metabolic changes.

PrgmNr 3539 - Investigation of matrisome genes within the druggable genome

View session detail

Author Block: **J. E. Brown**[1], M. Shuey[2], E. Gamazon[3], N. J. Cox[4]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Murfreesboro, TN, [3]VUMC Clare Hall, Univ. of Cambridge, Nashville, TN, [4]Vanderbilt Univ Med Ctr., Nashville, TN

Disclosure Block: **J.E. Brown:** None.

The functions of extracellular matrix-related genes are structural support and biochemical or biomechanical cues for cells or tissues. Based on results of studies of the phenotypic consequences of loss of genes that traffic fibrillar collagens to enable them to perform their roles in the extracellular matrix, we hypothesized that matrisome genes contribute to risk of neuropsychiatric disease in addition to other medical phenome. While extracellular matrix-related genes are a subgroup of all genes, this subgroup contains a wide variety that can be further broken down for comparisons. One such way to coalesce genes within the matrisome gene set is to focus on those within the druggable genome. Using the Illuminating the Druggable Genome (IDG) development levels, each gene was notated as Tdark (virtually unknown target with no known drug or small molecule activities), Tbio (target with no known drug or small molecule activities), Tchem (target has at least one ChEMBL compound), or Tclin (target has at least one approved drug). Within the matrisome gene set, we have identified 25 Tclin genes and 68 Tchem genes. Using PredixVU, the application of PrediXcan on Vanderbiltâ□□s biobank BioVU, we are able to identify associations between phenotypes and predicted gene expression for any given gene. Preliminary explorations were limited to Tclin matrisome genes within the 10k EUA PredixVU dataset. Genes with enriched phenotype sets include *SERPINC1* (increased gene expression associated with cancer and macular degeneration), *IL6* (increased gene expression associated with various cardiac issues), and *PLAT* (increased expression associated with secondary malignancies). Genes associated with neuropsychiatric disorders include *VEGFA*, *IL6*, *EGLN1*, and *IL23A*. Further investigations will use the 70k EUA PredixVU dataset and will expand to include Tchem matrisome genes. We will quantify this enriched phenome associated with these Tclin and Tchem matrisome genes using Mendelian randomization. This work will potentially allow for expanded use of the drugs and compounds associated with the matrisome genes within the druggable genome.

PrgmNr 3540 - Long-read amplicon sequencing of the polymorphic *CYP2D6* locus

View session detail

Author Block: **L. Zhu**, J. Harting, A. M. Wenger, Z. Kronenberg, P. Baybayan; Pacific BioSci.s, Menlo Park, CA
Disclosure Block: **L. Zhu:** Salary/Employment; Pacific Biosciences.
The *CYP2D6* gene affects the metabolism of ~25% of prescribed drugs in the United States. Therefore, it is necessary to identify the individual *CYP2D6* genotypes for the improvement of personalized medicine. However, current prevalent technologies, e.g., clinical microarray and qPCR, have limitations in genotyping the ~8 kb region at the highly variable *CYP2D6* locus. This study presents a simple long-range PCR method combined with PacBio HiFi sequencing for identifying duplications, deletions, polymorphisms, and gene conversions of the *CYP2D6* gene.
A two-step barcoded PCR approach was developed, including a first-step long PCR using gene-specific primers tailed with M13 sequences and a second-step amplification with barcoded M13 primers. This approach has been applied on 22 Coriell pharmacogenomics reference samples and 41 human saliva samples. Barcoded amplicons were pooled together for SMRTbell library preparation, sequenced on the PacBio Sequel II and IIe Systems. HiFi reads generated from PacBio sequencing were analyzed by the pbaa (PacBio amplicon analysis) pipeline. Nearly all (>99%) HiFi reads were on target for *CYP2D6*. Our results demonstrate that the long-range PCR in this study is robust and reliable to genotype real-world human samples. Furthermore, the pbaa analysis can process long amplicon HiFi reads efficiently. Overall, the sequencing results demonstrate that HiFi reads of 8-10 kb long with a median accuracy >99.9%, provide base-level resolution for identifying polymorphisms, indels, and gene conversions in the highly variable gene region. This study indicates the potential use of PacBio HiFi sequencing in clinical research applications.

PrgmNr 3541 - Targeting Alzheimerâ□□s Disease drug development via network analysis of GWAS data

View session detail

Author Block: **M. Mews**[1], B. Kunkle[2], L. Dumitrescu[3], T. J. Hohman[3], W. S. Bush[4]; [1]Systems Biology and Bioinformatics, Dept. of Nutrition, Case Western Reserve Univ., Cleveland, OH, [2]Univ. of Miami, Miami, FL, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Case Western Reserve Univ., Cleveland, OH
Disclosure Block: **M. Mews:** None.
Alzheimerâ□□s disease (AD) is the most common neurodegenerative disorder, affecting more than 30 million patients globally, but currently there exists a severe lack of effective drug treatments for AD even after 30 years of research; the FDA has just approved the first AD-drug aducanumab, which targets beta-amyloid, but its clinical effectiveness is still under heavy scrutiny. In our study, we focused on identifying potentially novel AD drug targets using a network-based approach that leverages existing, curated knowledge. We focused our analysis on 25 seed genes identified and validated as associated with AD in the largest AD meta-analysis to date by Kunkle *et al.* 2019. Subsequently, we leveraged existing high-confidence experimental and curated database entries from StringDB to identify functionally interacting partners of these seed genes. We further restricted our analysis to gene partners classified as known or predicted druggable gene targets according to the Drug Gene Interaction database (DGIdb). We excluded gene partners that were themselves seed genes and also four known early-onset AD genes (*PSEN1/2*, *APP*, and *MAPT*). These gene partner sets were subsequently analyzed in the context of the druggable genome via the generalized gene-set analysis tool MAGMA using two primary study sources of GWAS summary statisticsâ□□the previously mentioned Stage 1 results of the AD meta-analysis for AD risk (N = 63,926) and four analyses of a dataset derived from studying cognitive and global resilience to Alzheimerâ□□s disease by Dumitrescu *et al.* 2020 filtered to include either all participants (N = 5,108) or solely those with unimpaired cognition (N=3,820). These analyses test the hypothesis that variants within druggable gene partners may contribute to AD risk or protection. Two gene sets from the Stage 1 results analysis approached nominal significance, *CLU* gene partners (N = 76, p = 0.051) and *ADAMTS1* gene partners (N = 26, p = 0.055), whereas a single gene set, *PICALM* gene partners (N = 33, p = 0.031), was nominally significant when analyzing cognitive resilience in the context of cognitively unimpaired individuals. Driver gene partners associated with the *PICALM* gene set include known druggable genes *AAK1* (p = 0.017), *EGF* (p = 0.044), *GAK* (p = 0.007), and *IL7R* (p = 0.056) and the predicted druggable gene *TXNDC5* (p = 0.044). These findings suggest that future AD drug prioritization efforts may be more fruitful if researchers focus on druggable genes functionally adjacent to GWAS hits.

PrgmNr 3542 - Worse COVID-19 outcomes in patients with the hemoglobin beta sickle allele due to existing comorbidities and new acute events

View session detail

Author Block: **S-W. Luoh**[1], A. Verma[2], L. Gao[1], J. Minnier[1], The VA Million Veteran Program, S. Iyengar[3]; [1]VA Portland Hlth.Care System, Portland, OR, [2]Corporal Michael Crescenz VA Med. Ctr., Philadelphia, PA, [3]Louis Stokes Cleveland VA Med. Ctr., Cleveland, OH
Disclosure Block: **S. Luoh:** None.
Introduction. The sickle cell trait (SCT) in the hemoglobin beta gene (HbS; rs334) affects millions of Americans, especially African Americans (AA; minor allele frequency [MAF]=7.8%) and Hispanic Americans (HA; MAF=1%). We investigated the impact of SCT on the severity and sequelae of COVID-19 infection in the VA Million Veteran Program (MVP). Pre-COVID diseases and laboratory findings present in electronic health records (EHR), as well as acute events following 60 days of COVID-19 infection, and their mediating effect on COVID-19 mortality among SCT patients were examined. **Methods**. Pre- and post- COVID-19 clinical data on genotyped MVP participants (SCT+ = 2,729, SCT-= 129,848; COVID+=13,841, COVID-=118,736) was extracted from EHR. Outcomes analyzed were: severe disease (or mortality) vs. not severe (or survival). Ethnic-specific firth logistic regression for SCT was performed on European (EA), African (AA), Hispanic (HA) and Asian (ASA) groups, adjusting for sex, age, age2, and 20 genetic principal components. Ethnic-specific phenome-wide association (PheWAS and LabWAS) for SCT captured 20+ years of comorbidities and historical laboratory values and was used to contrast effects of COVID-19. Multiple testing corrections were applied. **Results**. HbS is associated with increased COVID-19 mortality in AA (N=3,749; OR=1.8 [1.14-2.84], p=0.01) with a similar trend in HA. PheWAS revealed significant associations of rs334 with phecodes for pulmonary embolism, chronic renal disease, diabetic kidney disease, hypertensive renal disease, gout, sickle cell disease/trait, and hemolytic anemia (FDR p Conclusions. SCT is associated with an elevated risk of mortality with COVID-19 infection. Both pre-existing chronic medical conditions and new acute events after COVID-19 may contribute to adverse COVID-19 outcomes with SCT.

PrgmNr 3543 - A Single-Assay Solution for Expanded Carrier Screening Relieves Existing Workflow Constraints and Provides More Comprehensive Analysis

View session detail

Author Block: **S. Statt**, J. Kemppainen, M. Church, C. Edwards, J. Turner, J. Janovsky, S. Filipovic-Sadic, J. R. Thibert, B. Hall, L. Chen, T. Markulin, J. Hernandez, J. Wisotsky, B. C. Haynes, G. J. Latham; Asuragen, Austin, TX
Disclosure Block: **S. Statt:** Major Stockholder/Ownership Interest; Asuragen.
More than a million individuals have been tested using Expanded Carrier Screening (ECS) since the first commercial assay was launched in 2009. Conventional methods like short-read sequencing either miss or poorly cover many genes or gene regions; a recent study found that 20.4% of pathogenic/likely pathogenic variants were â□□technically challengingâ□□ by NGS. Consequently, analysis of problematic and high-prevalence carrier screening genes requires specialized non-NGS methods, creating inefficient workflows and subpar detection rates, limiting access to only high throughput sophisticated labs. We sought to create a mid-sized panel that combines routine NGS targets with non-NGS targets in a single workflow that can enable broad access to ECS assays. To do this, we paired novel long-range PCR with long-read nanopore sequencing and algorithms that offer more comprehensive yet streamlined ECS analysis.
Samples were initially evaluated using the most challenging genes from the mid-sized panel as a proof-of-concept, namely *FMR1*, *F8*, *CYP21A2*, *HBA1/2*, *GBA* and *SMN1*. Genomic DNA was amplified, barcoded, pooled, prepped by ligation sequencing kit (ONT) and ran on R9.4.1 flow cells (ONT) using either the MinION or Mk1C. A custom pipeline resolved CGG genotypes and AGG interruptions in *FMR1*. For other genes, a bespoke sequence deconvolution method reconciled paralogous sequences, where applicable, along with pathogenic variants. We accurately detected multiple classes of pathogenic variation across 24 Coriell cell-line and residual whole blood samples, including large structural variants (such as inversions), SNVs, short tandem repeats, INDELs and CNVs. *FMR1* CGG repeats were accurately categorized as normal, intermediate, premutation or full mutation; they also fell within reference AmplideX PCR/CE *FMR1* Kit (RUO) precision ranges by repeat count. AGG interruptions were concordant with independent analyses, and informed a broad range of intergenerational expansion risk. Complex *CYP21A2* gene conversions, pseudogene fusions, and *CYP21A2/CYP21P1* gene copy numbers were correctly determined using as few as 100 reads per allele. *SMN1* analysis identified copy numbers ranging from 0 to â□¥4 in agreement with orthogonal assay calls.
The data demonstrate feasibility for a single-platform, multiplexed panel workflow that can accurately resolve different classes of challenging variants, and scale to include more conventional carrier genes. This method has the potential to address real-world gaps in carrier screening by integrating a nanopore sequencer with tailored PCR reagents and software for use in diverse laboratory settings.

PrgmNr 3544 - Cord blood DNA methylation gestational age estimation applied to dried blood spot DNA methylation

View session detail

Author Block: **K. Polinski**[1], S. Robsinson[1], D. Putnick[1], W. Guan[2], S. Mumford[1], E. Yeung[1]; [1]NICHD/NIH, Bethesda, MD, [2]Univ Minnesota, Minneapolis, MN
Disclosure Block: **K. Polinski:** None.
Objective: DNA methylation (DNAm) â□□clocksâ□□ have been used to predict gestational age (GA) at delivery and are typically applied to cord blood DNAm. Yet, dried blood spots (DBS) are routinely collected for newborn screening and have utility for investigating epigenomic features. Therefore, we estimated GA based on cord blood DNAm from the EPIC 850K array and compared clocks on DBS DNAm in an independent dataset.
Methods: EAGeR is a preconception cohort that followed women attempting pregnancy and throughout pregnancy for women who conceived. Clinical GA was determined using a combination of ultrasound, fertility monitor information, last menstrual period, and clinical assessment. Cord blood DNAm data was available in 391 singletons. Array-wide associations between clinical GA and DNAm beta values were identified using multivariable robust linear regression in the training dataset (70%, n=274). False discovery rate (FDR) significant CpG probes were then added as predictors in an elastic net regression of GA in which alpha was set to 0.5 and lambda was chosen using 10-fold cross-validation on the training dataset. The resulting CpG probes (i.e. EAGeR clock) were used to estimate DNAm GA in EAGeR and subsequently in an independent cohort (Upstate KIDS). Upstate KIDS includes 855 singletons and twins with DBS DNAm from the EPIC array. GA was based on birth certificate estimates.
Results: The average GA at birth in EAGeR was 38.9 weeks (SD 1.53; Range 31.1-44.8). We identified 10328 CpG probes associated with GA with FDR significance (**Conclusion:** Despite few overlapping CpG probes (11/225, 5%) in GA clocks from different cord blood DNAm datasets, the correlation to clinical GA in an external sample using DBS was similar. The correlation with clinical GA in preterm deliveries was reduced, potentially indicating that differences in clinical GA and DNAm GA may be biologically relevant in terms of cellular maturity and may have potential implications for development outcomes.

PrgmNr 3545 - Developmental genomics of congenital limb malformations: further evidence for gene dosage effects

View session detail

Author Block: **R. DUAN**[1], H. Hijazi[2], Y. Bayram[3], E. Karaca[4], D. R. Bertola[5], S. Isikay[6], E. Y. Gulec[7], H. K. Eker[8], Y. Sahin[9], Z. Ocak[7], H. Du[1], J. M. Fatih[1], G. Akay[1], S. Gu[10], Baylor-Hopkins Center for Mendelian Genomics, S. N. Jhangiani[11], Z. Coban-Akdemir[1,12], D. Pehlivan[1], R. A. Gibbs[11], J. E. Posey[1], C. M. Carvalho[13], J. R. Lupski[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Johns Hopkins, Baltimore, MD, [3]King of Prussia, PA, [4]Sanford Imagenetics, Sioux Falls, SD, [5]Inst. da Crianca, Sao Paulo, Brazil, [6]Gaziantep Univ., Gaziantep, Turkey, [7]Kanuni Sultan Suleyman Training and Res. Hosp., Istanbul, Turkey, [8]Dr Faruk Sukan Maternity and Childrenâ□□s Hosp., Konya, Turkey, [9]Necip FazÄ±l City Hosp., Kahramanmaras, Turkey, [10]Hong Kong S.A.R., China, [11]Baylor Coll. Med., Houston, TX, [12]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [13]Pacific Northwest Res. Inst., Seattle, WA
Disclosure Block: **R. Duan:** None.

Congenital limb malformations represent a broad spectrum of intrauterine developmental perturbations involving the upper and/or lower extremities, which can occur as a deformation, disruption, or an isolated developmental event (i.e., malformation) or be a manifestation of a partial trait, i.e., endophenotype, of a Mendelian condition. Phenotypic manifestations of extremity anomalies can involve anterior-posterior (AP), dorsal-ventral (DV), or proximal-distal (PD) planes of the body axes of development and can be associated with complex genetic etiologies. Genetic heterogeneity, reduced penetrance and variable expressivity have all been described in families. Using family-based genomics and rare variant analyses, we applied exome sequencing (ES) combined with whole-genome array-based comparative genomic hybridization (aCGH) to investigate 14 families with limb defects. Studies in 8 out of 14 families revealed likely pathogenic single nucleotide variants (SNVs) or copy number variants (CNVs) at previously reported â□□disease associated lociâ□□: *BHLHA9*, *HOXD13, GLI3, WNT10B* and *NPR2*. Multi-locus pathogenic variation (MPV) was observed in one family seemingly driven by the absence of heterozygosity (AOH) resulting from identity-by-descent (IBD); total AOH = 89.1Mb in a child of parents from the same small village, but with no known consanguinity. Notably, breakpoint junction analyses for 2 novel pathogenic duplication CNVs of *BHLHA9* were demonstrated to be generated by *Alu/Alu*-mediated rearrangement (AAMR). Interestingly, homozygosity for *BHLHA9* duplication CNV was observed in association with a more severe limb malformation, the Gollop-Wolfgang Complex. We propose a gene dosage model at this locus potentially underlies both the reduced penetrance and the more severe phenotype observed with homozygous duplication. Further studies will explore whether genes acting in the Apical Ectodermal Ridge (AER) may be particularly vulnerable to stochastic fluctuations in expression during development in the distal proximal plane and whether such a â□□gene dosage expression modelâ□□ significantly contributes to limb anomalies.

PrgmNr 3546 - Potential additive effect of fetal HR*APOL1*genotype and micronutrient deficiencies

View session detail

Author Block: **W. Bruner**[1], R. Davis[1], C. L. Simpson[2]; [1]Universtiy of Tennessee Hlth.and Sci. Ctr., Memphis, TN, [2]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN
Disclosure Block: **W. Bruner:** None.
Background: Preeclampsia (PE) is a hypertensive disorder of pregnancy that is associated with micronutrient deficiencies. Fetuses carrying two *APOL1* high risk (HR) variants is also associated with PE risk. We hypothesized a potential additive effect between HR *APOL1* genotype status and nutritional deficiencies would place individuals at a higher risk of developing PE. We assessed PE-associated risk of micronutrient deficiencies based on fetal *APOL1* genotype in pregnant African American women.
Methods: We used data from the Conditions Affecting Neurocognitive Development and Learning in Early Childhood (CANDLE) project and performed analyses using multivariate logistic regression to examine the association of PE with 2nd and 3rd trimester plasma folate and 25-hydroxy vitamin D concentrations. Concentrations were dichotomized into high or low categories. Folate deficiency was defined as a concentration less than 6 ng/mL. 25-hydroxy vitamin D deficiency was defined as a concentration less than 20 ng/mL. Further analyses assessed whether the inclusion of fetal *APOL1* genotype status modified the micronutrient association with PE.
Results: Compared to pregnancies with fetal LR status and high 3rd trimester vitamin D levels (reference group), the risk for preeclampsia was elevated among fetal LR/low vitamin D (OR 1.88; 95% CI 0.83,4.22), fetal HR/high vitamin D levels (OR 2.43; 95% CI 0.71, 8.28), and was highest for pregnancies with fetal HR/low vitamin D levels (OR 5.41; 95% CI 1.87,15.66). When assessing folate, compared to pregnancies with fetal LR status and high 2nd trimester folate levels, the risk for preeclampsia was elevated among fetal LR/low folate (OR 2.29; 95% CI 0.75,7.04) and among fetal HR/high folate (OR 1.95; 95% CI 1.03, 3.71). We were not able to assess 3rd trimester folate deficiency and fetal HR genotype due to low sample size.
Conclusions: Our findings are suggestive of a potential additive effect between low 25-hydroxy vitamin D and fetal HR *APOL1* genotype. Similar potential additive effect was found between low folate and fetal HR genotype.

PrgmNr 3547 - A novel polygenic score for microglial activation enhances models of Alzheimer's disease neuropathology

View session detail

Author Block: **E. Tio**[1], M. Milic[1], P. L. De Jager[2], D. Felsky[1]; [1]The Ctr. for Addiction and Mental Hlth., Toronto, ON, Canada, [2]Columbia Univ Med Ctr, New York, NY
Disclosure Block: **E. Tio:** None.
Neuroinflammation and the activation of microglia are among the earliest changes in Alzheimerâ□□s disease (AD). However, lack of direct access to living brain tissue has impeded the development of biomarkers of microglial states. To solve this, genetic markers may offer a non-invasive route to estimating burden of microglial activation in living people.
We recently performed a genome-wide association study (GWAS) of morphologically activated microglia measured postmortem in two regions of human cortex (midfrontal and inferior temporal). Based on these GWAS summary statistics, we used the clumping and thresholding method (PRSice-2) to calculate novel microglial activation polygenic risk scores (micPRS) in imputed genotype data from *n*=939 participants from the Alzheimer's Disease Neuroimaging Initiative (ADNI). For comparison, we calculated standard AD PRSs using summary statistics from the International Genomics of Alzheimer's Project (IGAP). Linear models were used to test associations between each PRS and AD diagnosis, ADAS-Cog scores at baseline, and cerebrospinal fluid (CSF) and brain imaging (AV45 and AV1451 PET SUVR) measures of amyloid and tau, with appropriate covariates including genomic principal components.
Our top inferior temporal micPRSs showed significant positive associations with all phenotypes ($3.57x10^{-4}$ fdr -2). Conversely, our midfrontal micPRSs showed significant associations with only PET amyloid ($P_{fdr} = 9.14x10^{-3}$). We then compared models including both micPRSs and the IGAP-derived AD PRS against models with just the latter. Our inferior temporal micPRSs significantly contributed to models of cognition and both amyloid and tau measured with PET and tau measured in CSF ($8.93x10^{-4}$ fdr -2, $9.50x10^{-3}$ 2 -2). However, they did not significantly improve models of CSF amyloid over and above the baseline AD PRS. Our midfrontal micPRS improved only models of PET amyloid ($3.68x10^{-3}$ fdr -3, $1.37x10^{-2}$ 2 -2).
This PRS-based approach to estimating microglial activation burden in living people produced results which resemble the associations of microglial activation with AD pathology identified directly in postmortem tissue. We also found that the addition of micPRS to models including only the AD PRS significantly improved predictions of multiple AD biomarkers. However, we observed strong variability in PRS performance across SNP inclusion p-value thresholds. Thus, further evaluation is required to establish the reliability of a microglial PRS in predictive models of AD.

PrgmNr 3548 - A novel test for parent-of-origin effects in population samples leveraging multiple phenotypes

View session detail

Author Block: **T. Head**[1], M. P. Epstein[2]; [1]Dept. of Biostatistics and Bioinformatics, Emory Univ., Atlanta, GA, [2]Emory Univ., Atlanta, GA
Disclosure Block: **T. Head:** None.

There has been long-established interest in identifying variants within the human genome that exhibit parent-of-origin effects (POEs) wherein the effect of an allele on phenotype expression depends on the parental origin of that allele. POEs can arise from many different phenomena including genomic imprinting and have been documented for many complex traits such as type 2 diabetes and body mass index. Traditional tests for POEs required family data to determine parental origins of transmitted alleles for analysis. For genome-wide association studies, most of which are performed on unrelated individuals for which allelic parental origin is unknown, the study of POEs has required more sophisticated statistical methods that exploit genetic patterns that we anticipate observing if a POE exists at a given locus. For example, Hoggart et al. [PLoS Genetics: e1004508] recognized that the existence of a POE results in increased phenotypic variance among heterozygotes compared to homozygotes and developed a novel population-based test based upon this observation. Here, we extend this method for identifying POEs to accommodate multiple phenotypes, which are increasingly collected in GWAS given widespread evidence of pleiotropy. We show that, in the presence of a POE, the covariance patterns among multiple phenotypes will differ between homozygotes and heterozygotes. We can assess such differences using a method that tests whether the phenotype covariance matrix among homozygotes differs from the analogous covariance matrix for heterozygotes. Our method is robust to non-normality of phenotypes and can easily adjust for population stratification and other non-genetic confounders. We evaluated our method through simulation studies and observed appropriate type I error and power to detect POEs when they exist. We further note that our method is orthogonal to standard tests of association between a SNP and multiple phenotypes and thus provides complementary information that can be used to assess top findings. Finally, we applied our method to GWAS studies of quantitative traits related to type I diabetes, post-traumatic stress disorder, and facial morphology and identified several variants of suggestive significance for follow-up investigation.

PrgmNr 3549 - A powerful and resource-efficient pipeline for association analysis of large-scale whole-genome sequencing studies

View session detail

Author Block: **Z. Li**[1], X. Li[1], H. Zhou[1], S. M. Gaynor[1], M. S. Selvaraj[2], A. Pampana[2], J. I. Rotter[3], C. J. Willer[4], G. M. Peloso[5], P. Natarajan[2], X. Lin[1], TOPMed Lipids Working Group; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]CVD Prevention Ctr., Massachusetts Gen. Hosp., Boston, MA, [3]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [4]Univ. of Michigan, Ann Arbor, MI, [5]Boston Univ., Boston, MA
Disclosure Block: **Z. Li:** None.

Introduction

With the emergence of large-scale whole-genome sequencing (WGS) data, there is a pressing need to identify genetic components of complex traits. It is resource-demanding to perform association analysis of hundreds of millions of variants, especially for the rare variant analysis.

Methods

We propose a resource-efficient analysis pipeline, STAARpipeline, for phenotype-genotype association analyses of WGS data, including single variant analysis and variant-set analysis. The single variant analysis in STAARpipeline provides valid individual *P* values of variants given a MAF or MAC cut-off, for example, MAC â¥ 20. The variant-set analysis in STAARpipeline includes gene-centric analysis and non-gene-centric analysis of rare variants. The gene-centric coding analysis provides five genetic categories: putative loss of function (pLoF), missense, disruptive missense, pLoF and disruptive missense, and synonymous. The gene-centric noncoding analysis provides eight genetic categories: promoter or enhancer overlaid with CAGE or DHS sites, UTR, upstream, downstream, and noncoding RNA genes. To achieve this, we develop FAVORAnnotator to functionally annotates the variants using the FAVOR database and generate an annotated genotype file for the input of STAARpipeline. The non-gene-centric analysis includes sliding window analysis with fixed sizes and dynamic window analysis with data-adaptive sizes. STAARpipeline also provides analytical follow-up of dissecting association signals independent of known variants via conditional analysis. We applied the STAARpipeline to analyze four quantitative lipid traits (LDL-C, HDL-C, TG, and TC) in 30,138 samples from the NHLBI Trans-Omics for Precision Medicine program.

Results

All analyses scale well in computation time and memory. The computation cost for STAARpipeline of each trait on 30,138 related samples are as follows: 1.5 hours for 200 2.10 GHz computing cores with 6 Gb memory of single variant analysis for variants with MAC â¥ 20; 2.5 hours and 15 hours for 200 cores with 5 Gb memory of gene-centric coding and noncoding analysis, respectively; 11 hours for 200 cores with 11 Gb memory of sliding window analysis; 20 hours for 800 cores with 15 Gb memory of dynamic window analysis. STAARpipeline takes 24 hours to summarize these results and perform conditional analysis of significant findings using one core with 25 Gb memory. We discover several conditionally significant associations with lipids, including a novel finding of an intergenic region near *SLC22A3* associated with total cholesterol.

Summary

The STAARpipeline is a fast and scalable tool for association analysis of extensive WGS studies.

PrgmNr 3550 - A quantile integral linear model to quantify genetic effects on phenotypic variability

View session detail

Author Block: **J. Miao**[1], Y. Lin[2], Y. Wu[3], B. Zheng[4], L. L. Schmitz[4], J. M. Fletcher[4], Q. Lu[1]; [1]Univ. of Wisconsin-Madison, Madison, WI, [2]Baylor Coll. of Med., Houston, TX, [3]Univ. of WisconsinMadison, Madison, WI, [4]Univ. of Wisconsinâ□□Madison, Madison, WI
Disclosure Block: **J. Miao:** None.
Detecting genetic variants associated with the variance of complex traits, i.e. variance quantitative trait loci (vQTL), can provide crucial insights into the interplay between genes and environments (GxE) and how they jointly shape human phenotypes. Robust vQTL findings can be used to prioritize candidate variants in GxE analysis. The genome-wide summaries of genetic effects on phenotypic variability (vPGS) also has the potential to aggregate information across numerous genetic loci and improve both statistical power and biological interpretability of GxE studies. We introduce a novel statistical framework named QUAIL to estimate genetic effects on the variance of quantitative traits. Our approach directly addresses several limitations of current vQTL methods, including a lack of robustness to non-Gaussian phenotypes and confounding effects on both trait levels and trait variability. In addition, QUAIL can be applied to both categorical and continuous predictors. Through extensive simulations and analyses of real data, we demonstrate that QUAIL provides robust, powerful, and computationally efficient vQTL mapping. Applied to UK Biobank (N=375,791), QUAIL identified 11 novel vQTL for body mass index (BMI). Top vQTL findings showed substantial enrichment for interactions with physical activities and sedentary behavior. Followed up in three well-powered longitudinal cohorts (total N= 21,961), the vPGS based on QUAIL show superior predictive performance on both population-level and within-individual BMI variability compared to existing approaches. Overall, QUAIL is a unified framework to quantify genetic effects on the phenotypic variability at both single-variant and vPGS levels. It addresses critical limitations in existing approaches and may have broad applications in future gene-environment interaction studies.

PrgmNr 3551 - A resource-efficient tool for phenome-wide rare variant association analysis in large-scale whole-genome sequencing studies, with application to TOPMed metabolomics data

View session detail

Author Block: **X. Li**[1], Z. Li[1], E. V. Feofanova[2], C. Quick[1], H. Zhou[1], S. M. Gaynor[1], A. C. Morrison[2], E. Boerwinkle[2], J. I. Rotter[3], B. Yu[2], X. Lin[1], TOPMed Metabolomics and Proteomics Working Group; [1]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [2]The Univ. of Texas Hlth.Sci. Ctr. at Houston, Houston, TX, [3]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA
Disclosure Block: **X. Li:** None.

Introduction

Large-scale whole-genome sequencing (WGS) studies have enabled the analysis of rare variants (RVs) associated with complex human traits. Phenome-wide association study (PheWAS) of WGS data provides a unique opportunity to identify the pleiotropy effect of RVs on multiple traits. Due to the vast number of RVs in WGS studies, it is computationally expensive to apply the RV analysis pipeline to analyze a large number of traits, one at a time. Existing PheWAS pipelines are designed for genome-wide association studies and focus on single-variant analyses of common variants. Scalable PheWAS pipelines for RV association analysis of WGS data are currently lacking.

Methods

We propose STAARphewas, a powerful and resource-efficient tool for RV PheWAS in large-scale WGS studies. STAARphewas accounts for population structure and relatedness for both continuous and dichotomous traits by fitting the generalized linear mixed model using sparse genetic relatedness matrices. It then provides various strategies for grouping coding and non-coding variants based on functional categories, and incorporates multiple numerical functional annotations (for example, annotation principal component scores) to further increase power using the STAAR framework. To reduce the computation cost in PheWAS, we propose a new analytical strategy in STAARphewas that allows for data preprocessing and analysis of multiple traits simultaneously. STAARphewas only requires extracting the genotype and functional annotation data once when analyzing multiple traits, which substantially reduces the computation cost while maintaining exactly the same results compared to analyzing RV associations of each phenotype one at a time. In addition, STAARphewas provides conditional analyses to identify RV-set signals independent of nearby common variants.

Results

We applied STAARphewas to identify RV-sets associated with >100 quantitative metabolomics traits in ~12,000 related samples from the NHLBI Trans-Omics for Precision Medicine program Freeze 8 data. STAARphewas saves at least 10x computation time than existing methods, requiring 7 hours and 18 hours for 10 cores with 8 Gb memory per trait of gene-centric coding and noncoding analysis, respectively. We detected several genome-wide significant RV associations with metabolomics.

Conclusion

We propose STAARphewas as a resource-efficient framework for phenome-wide rare variant association analysis, while incorporating multiple variant functional annotations to further improve power. The STAARphewas pipeline provides an essential solution for RV PheWAS in large-scale WGS studies.

PrgmNr 3552 - An extended fine-mapping framework to identify causal genes from transcriptome-wide association studies (TWAS)

View session detail

Author Block: **S. Zhao**[1], W. Crouse[1], K. Luo[2], M. Stephens[1], X. He[1]; [1]Univ Chicago, Chicago, IL, [2]Chicago, IL

Disclosure Block: **S. Zhao:** None.

While GWAS have identified a large number of loci associated with complex traits, causal genes in these loci often remain unknown. TWAS propose to nominate candidate genes by testing associations of cis-genetic components of gene expression, inferred from expression QTL (eQTL) data, with the trait. While widely popular, TWAS may find associations in non-causal genes, when their eQTLs are shared with other nearby causal genes (co-regulated genes), or have pleiotropic effects on the phenotype. Methods have been proposed to better integrate eQTL and GWAS data. Colocalization methods test if gene expression and GWAS share the same causal variants. These methods usually assume a single causal variant for both gene and trait, are sensitive to parameters, and still cannot guarantee causality. Mendelian Randomization (MR) methods, on the other hand, use cis-eQTLs as instruments to test causality of gene expression on phenotype. MR, however, is difficult to apply in this setting because of the small number of instruments and their correlations due to linkage disequilibrium. We propose a strategy to identify causal genes from joint eQTL-GWAS analysis. Our method (causal-TWAS, cTWAS) is a regression model where response is phenotype, and covariates include cis-genetic components of expression of all nearby genes and genotypes of all nearby variants. By having expression and variants in the same model, cTWAS accounts for co-regulation among genes, and pleiotropic effects where variants directly act on the phenotype without affecting any gene. To make inference, we use a Bayesian variable selection approach with sparse priors for the gene and variant effects. cTWAS extends a fine-mapping method, SuSiE, to estimate these prior parameters from genome wide data and infers causal genes in any specific loci. We simulate using real genotype data from UK Biobank, under a realistic genetic architecture where gene expression explains a small percent of heritability. Under this scheme, existing methods (TWAS, SMR, coloc) all suffer from high false positive (FP) rates. In contrast, cTWAS produces calibrated FP rates while maintaining good power. We are now applying cTWAS to UKBB GWAS summary statistics. cTWAS is available in an efficient and easy-to-use R package.

PrgmNr 3553 - An iterative approach to adjustment for covariates in genetic association studies using GWAS summary statistics

View session detail

Author Block: **O. Vsevolozhskaya**[1], D. W. Fardo[1], X. Tong[2]; [1]Univ. of Kentucky, Lexington, KY, [2]The Natl. Inst. of Environmental Hlth.Sci., Durham, NC
Disclosure Block: **O. Vsevolozhskaya:** None.
Adjustment for covariates or correlated secondary traits in genome-wide association studies (GWASs) is common practice and has at least two purposes: (1) to account for potential confounding factors that can bias SNP effect estimates, and (2) to improve statistical power by reducing residual variance. Researchers routinely adjust for covariates such as age, gender, and the first dozen genetic principal components. Additional covariate adjustments may include heritable secondary outcomes with the motivation of identifying the direct genetic effect of the variant on the primary outcome of interest. For example, one may be interested in discovery of SNPs directly influencing insulin resistance while adjusting for an individual's body mass index. This adjustment for the secondary outcome is warrant if a genetic variant is associated with both the primary outcome and the covariate used for adjustment. However, if a genetic variant only influences the covariate and not the primary outcome, adjusting for the covariate induces a spurious association between the genetic variant and the primary outcome. In practice, when performing GWASs for a cohort with comprehensive genotyping and phenotyping (e.g., in electronic health record or phenomewide association study), it is difficult to exhaust all reasonable covariate adjustments without leading to an unintended bias introduced with respect to the primary outcome. It is also impractical to assume a single desired model across all researchers querying the resulting database of results. Thus, statistical approaches that can adjust for covariates based on reports of marginal associations or recover marginal associations from reports of covariate-adjusted associations, without having to re-run GWAS, are of interest. Here, we detail a statistical procedure that allows a researcher to perform these tasks back and forth almost instantaneously, given access to the original data. In case one does not have access to the original data but only the GWAS summary statistics, we provide theory of and simulation study results of approximated adjustments and marginalization.

PrgmNr 3554 - Ancestry-matching cases to controls from the GLAD database to accelerate genetic research in under-represented populations

View session detail

Author Block: **D. M. Veliz-Otani**, D. Loesch, B. Mitchell, S. J. Kittner, T. O'Connor; Univ. of Maryland Baltimore, Baltimore, MD
Disclosure Block: **D.M. Veliz-Otani:** None.
Background. Latino populations are under-represented in GWAS and use of the available samples is constrained by data privacy concerns. Development of a platform that provides allele frequencies from ancestry-matched controls to generate summary association statistics would contribute to closing this diversity gap by reducing the recruitment burden. We tested the performance of a proof-of-principle procedure that matches cases to controls from Genetics of Latin American Diversity (GLAD) database, a large collection of publicly available Latin American genomes. **Methods.** We obtained genotypes of 942 ischemic stroke cases from the Stroke Genetics Network (SiGN) and 18,102 potential controls with appropriate informed consent from GLAD. All subjects self-identified as Latino/Hispanic. Genotypes of SiGN samples were imputed through the Michigan Imputation Server-1.2.4. and ~9M SNPs remained after QC. Principal components were estimated by PCAiR and kinship coefficients by PCRelate. The final sample size was 927 cases and 12,956 potential controls after filtering out related subjects. We did a greedy n:1 control:case matching (1â¤ n â¤4) based on the Euclidean distance of the first 10 PCs weighted by their eigenvalues. Association was tested by a generalized linear mixed model using GENESIS, both with and without PCs as covariates. All models included sex, and age (for controls) or age at stroke (for cases). In a separate run, we added as covariates the phenotypes used as the ascertainment basis for each GLAD cohort (diabetes, asthma and hemophilia). If an ascertainment variable was unavailable for a cohort, we assumed the condition was absent in subjects from such cohort. **Results.** The distribution of the first three PCs of cases and matched controls was similar, for all values of 1 â¤n â¤4. Results of GWAS with and without PCs as covariates were virtually identical. The genomic control inflation factor (λ_{GC}) was 1.086 for the 4:1 GWAS without PCs nor ascertainment phenotypes as covariates. This model replicated only one known risk gene (*C10orf143*), but identified 16 previously unreported loci, likely false signals due to ascertainment bias or batch effect. The model including ascertainment phenotypes as covariates had a λ_{GC} value of 0.976, and identified 3 loci, none previously associated with ischemic stroke. **Conclusion.** Ancestry-matching by Weighted Euclidean distance on the PCA space reduces biases due to population stratification, but fails to adjust for other sources of confounding, such as ascertainment bias or batch effects. A platform that generates summary statistics of matched controls must account for other sources of confounding.

PrgmNr 3555 - Bayesian estimation of Allele-Specific Expression (ASE) in the presence of phasing uncertainty for identifying genes with aberrant expression

View session detail

Author Block: **X. Zou**[1], K. Patterson[2], University of Washington Center for Mendelian Genomics, A. Iyengar[1], P. S. Kishnani[1], G. E. Crawford[1], T. E. Reddy[1], A. S. Allen[1], W. H. Majoros[1]; [1]Duke Univ., Durham, NC, [2]Univ of Washington, Seattle, WA

Disclosure Block: **X. Zou:** None.

Many genetic causes of the rare disease involve changes in gene expression. While noncoding variants are difficult to interpret, allele-specific expression (ASE) is one indicator of a possible genetic regulatory effect. Current methods for detecting ASE are hampered by the difficulty of accurately phasing heterozygous sites within a gene. We describe a novel Bayesian hierarchical model, BEASTIE, that circumvents this problem by jointly modeling ASE and phasing so that they can mutually inform each other. We show that BEASTIE is more accurate than existing approaches that use a single variant or assume a single phase. In addition, we illustrate BEASTIEâ□□s utility in a clinical setting by analyzing 6 glycogen storage disease (GSD) patients who lack a genetic explanation of their disease after routine clinical sequencing. We identify one patient with significant ASE in AGL expression, a known cause of debrancher enzyme deficiency, in patients GSDIIIa.

PrgmNr 3557 - Consistent summary counts based rare variant analysis for discovery of germline predisposition genes using public genotype summary data as controls

View session detail

Author Block: **W. Chen**[1], G. Wu[2]; [1]St Jude Children's Res. Hosp., Memphis, TN, [2]262 Danny Thomas Place, Memphis, TN
Disclosure Block: **W. Chen:** None.
It is common that sequencing resources are invested on cases with few or no matched controls of healthy individuals, which hinders the discovery of germline predisposition genes through case-control association tests. To circumvent this problem, genotype summary counts from public datasets, such as gnomAD, can serve as convenient controls for rare variant association test. However, systematic inflation and false positives can arise if confounding factors, such as pipeline differences and population structure, are not well addressed. We propose a framework for Consistent summary Counts based Rare Variant burden test (CoCoRV) to address these challenges. It consists of several novel elements: consistent variant processing allowing state-of-the-art variant QC, association test accounting for different ethnic groups, improved association test inflation estimation for better assessment of systematic inflation, and a novel method to detect variants in high linkage disequilibrium using summary allele counts which is critical for removing false positives. When applied to pediatric brain tumor and acute lymphocytic leukemia, top genes include known cancer predisposition genes and an interesting new candidate gene. When applied to the brain tumor patients in The Cancer Genome Atlas (TCGA), we discovered and statistically validated two genes associated with brain tumor. One was TP53, a known cancer predisposition gene but missed by a previous TCGA germline analysis and the other gene *ABCB8* is likely a novel cancer risk gene. Given that potential confounding factors are well controlled after applying the framework, our tool provides a highly cost-effective solution to discovering risk genes enriched with pathogenic rare variants.

PrgmNr 3558 - DeepRsq: an improved imputation quality metric via a machine learning method

View session detail

Author Block: **Q. Sun**[1], Y. Yang[2], C. Fuchsberger[3], Y. Li[4]; [1]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]Sun Yat-sen Univ., Guangzhou, China, [3]Eurac Res., Bozen, Italy, [4]Univ North Carolina, Chapel Hill, NC

Disclosure Block: **Q. Sun:** None.

Background: Genotype imputation is an approach broadly used in genetics studies to estimate missing genotypes with the aid of a reference panel, but not all variants in the reference panel can be well imputed. As an estimate of the true imputation quality (true R2), researchers rely on variant-level statistics provided by the various imputation tools. However, these standard quality metrics (referred to hereafter as Rsq) have been found less informative for low frequency (minor allele frequency (MAF) Methods: We propose deepRsq, a novel machine learning enhanced quality metric by using an eXtreme Gradient tree Boosting (XGBoost) algorithm to effectively incorporate information from other variant-level features. Specifically, deepRsq considers a total of 81 variant-level features, including 11 population genetics statistics (e.g., Fst) from six major continental populations, minor allele counts (MAC) and MAFs in five major continental groups estimated from the Trans-Omics for Precision Medicine (TOPMed) cohorts, the 2nd - 4th moments of the five MACs, as well as the estimated MAF and Rsq in imputation target cohort. We divide the variants into three categories: common (MAF > 5%), low frequency (MAF in [0.5%, 5%]) and rare (MAF Results: We applied deepRsq to the Cystic Fibrosis cohort (n=5095) with whole genome sequencing data available, which serves as the truth for evaluating the performance of the two quality metrics. We found that deepRsq leads to much improved performance. For example, deepRsq increases the squared Pearson correlation with true R2 by 15.3% - 47.9% and decreases the root mean square error by 20.1% - 49.5%. We further applied deepRsq to UK Biobank European ancestry participants with whole exom sequencing (WES) data (n=99,990) where imputation was performed using genotyping array markers and WES data were used for evaluation purpose. We found that deepRsq rescues a larger number of decently-imputed signals that would have been filtered out by Rsq, and more effectively removes badly-imputed variants than Rsq. We further assessed deepRsqâ□□s performance in six additional multi-ethnic cohorts, observing similarly promising improvement. **Conclusion**: Our extensive evaluations demonstrate that deepRsq outperforms Rsq for post-imputation quality control. We anticipate that deepRsq will be valuable for future genotype imputation tasks.

PrgmNr 3559 - ExPheWas: a browser for gene-based pheWAS associations

View session detail

Author Block: **M-A. Legault**[1,2], L-P. Lemieux Perreault[2], M-P. Dube[2,1]; [1]UniversitÃ© de MontrÃ©al, Montreal, QC, Canada, [2]Montreal Heart Inst., Montreal, QC, Canada
Disclosure Block: **M. Legault:** None.
The identification and characterization of genetic associations with human phenotypes can improve our understanding of disease etiology and support the discovery of novel therapeutic targets.
The UK Biobank is an excellent resource for such research. It includes >500,000 participants with available genotypes and linkage to health records including cancer, hospitalization and death registries. We used this cohort to conduct a phenome-wide association study (pheWAS) of all protein coding loci. We implemented a gene-based principal component analysis approach based on common genetic variants from 19,114 protein coding regions and tested their association with 1,210 phenotypes including anthropometric measurements, laboratory biomarkers, algorithmically-defined cardiovascular outcomes and health records. The results of this analysis are publicly available in a user-friendly browser https://exphewas.statgen.org.
As a proof of concept, we characterized the 137 genes associated with atrial fibrillation at a false discovery rate of 1%. Using enrichment analysis, the genes identified were strongly enriched for relevant biological processes such as cardiac muscle contraction (P_{adj}=9.5 Ã 10^{-6}) and antiarrhythmic drug targets in the ChEMBL database. By further investigating possibly novel genes, we prioritized MYOT as a likely atrial fibrillation gene. In ExPheWas, this gene is strongly associated with heart rate (P=8.6 Ã 10^{-31}) and atrial fibrillation (P=4.9 Ã 10^{-11}). MYOT is a component of sarcomeric Z-disks and mendelian mutations of this protein are associated with cardiomyopathy.
ExPheWas is a database of gene to phenotype associations that may be used for follow-up of genetic studies and for enrichment analysis.

PrgmNr 3560 - Gene-based association approach in family samples using GWAS summary statistics

View session detail

Author Block: **X. Xu**[1], P. Wang[2], N. Liu[3]; [1]Indiana Univ. at Bloomington, Bloomington, IN, [2]Bloomington, IN, [3]Indiana Univ., Bloomington, IN
Disclosure Block: **X. Xu:** None.
Genome-wide association studies (GWASs) have led to rapid growth in detecting associations between diverse phenotypes and genes. Owing to a great number of GWAS summary statistics publicly accessible nowadays, plenty of gene-based association tests have been proposed to require GWAS summary statistics only rather than utilizing individual-level genotypes. However, these adaptive association tests are limited to unrelated individuals. Challenges in performing gene-based association tests with summary statistics still exist in family samples. Meanwhile, the linear mixed model has been increasingly expanded in GWAS to handle the relatedness in the samples because of its flexibility and effectiveness. Still, how to perform gene-based association tests with GWAS summary statistics estimated from the linear mixed model remains unknown. To address these two challenges, in this study, we show that the correlation matrix between the marginal Z scores, which is estimated by the linear mixed model, can be approximated by the linkage disequilibrium matrix of the associated SNPs due to the block diagonal structure of the kinship matrix. Thus, the existing summary data-based methods for unrelated individuals can be directly applied to family samples without any modifications. Simulation results further demonstrate that the proposed procedure could also well control the type I error rate in a variety of situations. In a practical setting, we exemplify the proposed approach with dental GWAS data.

PrgmNr 3561 - Gene-burden tests, gene-environment interactions and time-to-event data analysis within an efficient whole genome regression framework for large-scale biobanks

View session detail

Author Block: **J. Mbatchou**[1], Y. Zou[2], A. Ziyatdinov[3], J. L. Marchini[4], Regeneron Genetics Center; [1]Regeneron Genetics Ctr., TARRYTOWN, NY, [2]Univ. of Chicago, Chicago, IL, [3]Regeneron Genetics Ctr., Regeneron Pharmaceuticals, Tarrytown, NY, [4]Regeneron Genetics Ctr., Tarrytown, NY
Disclosure Block: **J. Mbatchou:** Major Stockholder/Ownership Interest; Regeneron Pharmaceuticals. Salary/Employment; Regeneron Pharmaceuticals.
The past decade has seen an unprecedented rise in the amount of phenotypic data available through the use of electronic health records and self-reported information. Large-scale biobanks have provided unique opportunities for researchers to make novel findings as well as validate existing targets and discover new indications for existing therapies. These biobanks provide rich phenotypic information, including time-to-event (TTE) data which not only inform on disease status but also incorporate age at onset information, which can provide better power in association analyses compared to using standard binary outcomes. They also constitute a rich data resource to explore and identify gene by environment (GxE) or gene by gene (GxG) epistatic effects, which require environmental exposure information as well as larger sample sizes for sufficient power. We have previously proposed REGENIE as an efficient computational method to analyze both quantitative and binary traits (including highly imbalanced traits) in large-scale biobanks that can highly reduce the computation time, all the more when thousands of phenotypes are to be analyzed, while accounting for population structure and relatedness. We have added three major new features to the REGENIE method 1) a flexible set of gene-based tests that utilize variant annotation files which are ideally suited for the analysis of the UK Biobank Exome sequencing dataset; 2) flexible GxE and GxG interaction tests for quantitative and binary traits, where we use robust standard errors and heteroscedastic linear models to prevent from inflation due to violations in homoscedastic assumption for quantitative traits; 3) an efficient approach for the analysis of TTE data which shares similar computational benefits as the REGENIE approach for binary traits and can handle highly censored phenotypes through a fast Firth-based association test within a Cox regression model framework. We demonstrate the performance of the proposed features through simulations and real data application in UK Biobank with up to ~400,000 samples. We also highlight the computational benefits of our method when run on multiple TTE phenotypes compared to several existing methods as well as contrast its performance against a case-control analysis approach within a logistic regression framework.

PrgmNr 3562 - Gene-environment interaction testing using machine learning approaches and robust test statistics: a sex-specific risk score component for lung function

View session detail

Author Block: **J. Hecker**[1,2], D. Prokopenko[3,2], M. Moll[1,4,2], W. Kim[1,2], D. Qiao[1,2], B. D. Hobbs[1,4,2], M. H. Cho[1,4,2], S. T. Weiss[1,2], D. L. DeMeo[1,2], C. Lange[5,1]; [1]Channing Div. of Network Med., Brigham and Women's Hosp., Boston, MA, [2]Harvard Med. Sch., Boston, MA, [3]Genetics and Aging Unit and McCance Ctr. for Brain Hlth., Massachusetts Gen. Hosp., Boston, MA, [4]Div. of Pulmonary and Critical Care Med., Brigham and Womenâ□□s Hosp., Boston, MA, [5]Dept. of Biostatistics, Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA

Disclosure Block: **J. Hecker:** None.

Rationale: Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with complex traits. However, the effect of a genetic variant on a complex trait can be influenced/modified by environmental exposures. The identification and understanding of such gene-environment interactions can provide new insights into the pathways and mechanisms. Existing methods for gene-environment interaction testing consider single variants or polygenic risk scores, suffering from low statistical power or limited biological interpretation. Furthermore, misspecified environmental main effects can invalidate interaction results. We propose a flexible framework for gene-environment interaction testing for quantitative traits that can adaptively screen for interactions while providing robustness against misspecifications.

Methods: Our approach is based on sample splitting and the utilization of so-called multiply-robust test statistics. Sample splitting allows to screen for potential interactions between multiple SNPs and environmental factors, combine signals, and test them in different parts of the data. The design of the test statistics provides robustness against misspecifications of environmental main effects and enables alternating the roles of screening and testing datasets, restoring efficiency. One important example is testing for interaction between an environmental factor and polygenic risk score SNPs.

Results: In extensive simulation studies, we demonstrate that our approach controls the type 1 error rate in a wide range of scenarios, including population stratification, gene-environment correlation, misspecified environmental main effects, heteroscedasticity, and non-normal error distributions. We also applied our approach to lung function data in the UK Biobank. For FEV_1, FVC, and FEV_1/FVC, we analyzed gene-environmental interactions between previously identified GWAS hits and multiple environmental factors, including sex, pack-years of smoking, and current smoking. For all three traits, we observed a highly significant interaction with sex (pConclusions: We propose a robust and flexible framework for gene-environment interaction testing for quantitative traits. In an application to UK Biobank data, we identified sex-specific associations between sets of SNPs and lung function.

Funding: National Heart, Lung, and Blood Institute [K08HL136928, U01HL089856, U01HL089897, P01HL120839, P01HL132825]. National Human Genome Research Institute [R01HG008976, 2U01HG008685]. This research was conducted by using the UK Biobank resource under application number 20915.

PrgmNr 3563 - HaploSoup: IBD-based encoding for improved PCA genome matching

View session detail

Author Block: **R. Laboulaye**[1], V. Borda[2], B. Guo[3], D. Loesch[4], T. D. O'Connor[1]; [1]Univ. of Maryland Sch. of Med., Baltimore, MD, [2]Univ. of Maryland Sch. of Med., Petropolis, Brazil, [34]Univ. of Maryland Baltimore, Baltimore, MD
Disclosure Block: **R. Laboulaye:** None.
Background: The advent of biobanks enables leveraging existing genomic data for use as controls for GWAS or admixture mapping studies. This is particularly useful for underrepresented populations such as Latin Americans whose studies often have lower sample sizes than comparable European studies. Retrieving suitable controls from a biobank requires a genome matching approach that can adequately differentiate individuals by population, a difficult challenge for Latin American populations where continental admixture can dominate fine-scale variation in Native American ancestry. While this is typically done by searching in a principal component space generated directly on genomic variants, we develop an IBD-based encoding that is applied prior to PCA to create a search space that better separates fine-scale populations and yields better matching results.
Methods: Population structure can be inferred through patterns of haplotype sharing. While LD pruning yields variants that each represent a larger haplotype, when applied to data with a variety of admixture proportions, it can erase haplotypes that constitute fine-scale structure. As an alternative, we propose HaploSoup, an IBD-based encoding that allows PCA to operate directly in haplotype segment space. Given a set of samples, we use the Positional Burrows Wheeler Transform to align our sequences and find all maximal length segment matches. We then apply a greedy reconstruction criteria to iteratively select representative segments to add to our encoding. The segments are scored according to their length and the as yet uncovered portion of the data that they cover. The selected set of segments is used to transform our samples from binary vectors of variants to binary vectors of segments, which can then be reduced with PCA.
Results: We evaluate our encoding approachâ□□s ability to improve genome matching for Latin Americans using both simulated and 1000 Genomes Project data. The evaluation uses a subsampled population as a query and measures how many of the matched genomes come from the correct population. We use principal components generated on genomic variants as a baseline and perform matching using a variety of distance metrics. We simulate admixed American populations with varying admixture proportions and find that HaploSoup improves matching accuracy from 74.2% to 86.9%. We then use the Latin American 1000 Genomes populations, MXL, CLM, PUR, and PEL, to repeat our experiment and find that our encoding improves matching accuracy from 53.6% to 66.9%. The Julia implementation of HaploSoup is highly parallel and can easily be added as a preprocessing step to existing matching pipelines.

PrgmNr 3564 - High-dimensional regression of continuous secondary traits under extreme phenotype sampling

View session detail

Author Block: **L. Ford**[1], H. Shen[2], H-W. Deng[3], C. Xu[1]; [1]Univ. of Oklahoma Hlth.Sci. Ctr., Oklahoma City, OK, [2]Tulane Univ, New Orleans, LA, [3]Tulane Univ., New Orleans, LA
Disclosure Block: **L. Ford:** None.
Introduction: To reduce high costs associated with genotype sampling, statistical methods have been developed to improve efficiency of parameter estimation on smaller samples. Extreme phenotype sampling has been shown to improve statistical power by using only extremes of the primary phenotype distribution as compared to random sampling and case-control methods. Secondary phenotype traits can also be evaluated after sampling extremes of the primary trait to further reduce cost. Maximum likelihood methods have already been developed to evaluate continuous secondary traits under extreme phenotype sampling. However, these likelihood approaches are unreliable under high-dimensional regression (i.e., joint modeling of genetic factors or a large number of covariates with a relatively small number of observations). Our objective is to develop a novel method than can address the high-dimensional problem in secondary trait analysis.
Methods: In our preliminary study, we conducted simulations with randomly generated data. The generated data included primary phenotype, secondary phenotype, and gene expression information. Data were generated for 1,000 individuals of which 200 were sampled for analysis by extreme phenotype sampling. Previously developed maximum likelihood estimation models were applied to determine the association between the secondary phenotype and a primary gene expression of interest. The data was generated in a way that the primary gene expression of interest, in truth, had no association with the secondary phenotype trait. From 1,000 simulations per scenario, we calculated the type I error (at an alpha level of 0.05) as the number of genes increased. We considered scenarios with 10, 100, 500, and 1000 gene expressions.
Results: We found that increasing the number of genes in the model led to an increase in the type I error. The type I error was 5.4% for 10 genes, 6.6% for 100 genes, and 16.3% for 500 genes. With 1,000 genes in the model the likelihood function was non-finite and could not be maximized.
Conclusion: The current likelihood methods for analyzing secondary traits under extreme phenotype sampling cannot be used to handle large numbers of factors jointly, like gene expressions. This limitation motivates the development of high-dimensional methods to analyze such data. This development could lead to improved statistical power and reduced cost in gene association studies.

PrgmNr 3565 - Human Genetics Structure Activity Relationship (HG-SAR): A novel methodology to link human genotype-phenotype associations to high throughput functional activity data and validate targets for human disease

View session detail

Author Block: **M. R. Miller**[1], X. Chen[1], H. Kim[1], A. Mancini[2], S. Stephan[3], X. Leroy[3], C. Normand[2], C. L. Hyde[4], J-P. Fortin[1], E. Fauman[1]; [1]Pfizer, Cambridge, MA, [2]Domain Therapeutics, Montreal, QC, Canada, [3]Domain Therapeutics, Strasbourg, France, [4]Pfizer, Groton, CT
Disclosure Block: **M.R. Miller:** Salary/Employment; Pfizer, Inc.
Large scale efforts to perform exome and whole genome sequencing have identified millions of novel genetic variants and hundreds of new genetic associations. Exome-sequencing specifically can help us to pinpoint the causal gene for an association by identifying coding mutations associated with a trait or disease. However, many of the variant-phenotype associations identified by exome sequencing are with missense variants with unknown function, leaving the questions of the direction and strength of the effect unanswered. In order to fully exploit the power of naturally occurring mutations, we present a novel target validation approach: *Human Genetics Structure Activity Relationship* (HG-SAR). This method leverages the power of human genetics by linking the molecular, cellular, and clinical consequences of rare missense variants to accelerate the discovery of novel disease relevant targets and pathways. Focusing on the GPCR family, we functionally characterized over 2400 missense mutations across 42 GPCRs. All the variants selected for functional characterization were identified in at least one subject in the UK Biobank, allowing us to correlate the variant in-vitro function with phenotypes of interest. Follow-up HG-SAR analyses allow us to regress the variant-phenotype data in a gene against the functional read-out for each variant, which can help to establish whether a gene is causal for a trait/disease, as well as establish directionality and the geneâ□□s maximal potential therapeutic impact. As a positive control, we have applied this method to MC4Rand demonstrated a linear relationship between impaired MC4R G_{15} signaling and body weight. For each 10% decrease in MC4R G_{15} signaling, bodyweight increases by 0.76 kg. We next applied this method to GIPR, a gene with known GWAS associations with BMI for which the direction of effect was unclear. HG-SAR established that loss of GIPR signaling activity is associated with decreased body weight, establishing the directionality for modulating this target. This novel human target validation strategy, likely applicable across all protein families, is an exciting new method to identify and validate potential drug targets.

PrgmNr 3566 - Identity-by-descent mapping in biobank-scale datasets

View session detail

Author Block: **L. E. Petty**[1], R. Bohlender[2], H-H. Chen[3], J. Baker[4], G. Evans[5], J. E. Below[1], C. D. Huff[6]; [1]Vanderbilt Genetics Inst., Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]1155 Pressler Street, Houston, TX, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Nashville, TN, [5]NASHVILLE, TN, [6]U.T. MD Anderson Cancer Ctr, Houston, TX

Disclosure Block: **L.E. Petty:** None.

Identity-by-descent (IBD) mapping is a gene mapping approach that utilizes IBD segments to identify loci that are enriched for sharing in disease cases compared to controls. This approach allows for detection of loci that harbor variants that are too low frequency for imputation to perform well, which may have a greater penetrance than common variants well-powered for genome-wide association studies (GWAS).

We have developed a tool for performing biobank-scale IBD mapping, IBDMap. We leverage IBD segment data, detected using an external segment detection tool, and perform permutation-based testing to determine enrichment of IBD sharing by comparing rates of sharing in case-case pairs to rates in case-control pairs at each segment breakpoint, genome wide. Calculations for each breakpoint are performed independently and parallelized via a process queue. A complementary Python package performs map-reduce on multiple rounds of permutations, for computational efficiency in large-scale datasets. IBDMap performs flexible multiple testing correction, utilizing a permutation-based family-wise error or false discovery rate approach to account for the correlation structure. We also introduce a new haplotype-based approach which filters IBD segments that are unlikely to harbor rare, pathogenic variants due to high population frequency.

We applied IBDMap to a range of cardiovascular phenotypes in BioVU, Vanderbiltâ□□s DNA biobank with linked electronic health records. Using diagnostic billing codes, we determined case status for 69,819 European ancestry individuals genotyped on the MEGA^{EX} array for myocardial infarction, atherosclerosis, dyslipidemia, dilated cardiomyopathy, and atrial fibrillation. We then applied IBDMap for each phenotype to consensus IBD segments detected using iLASH, hap-ibd, and GERMLINE. We identified regions on chromosome 2 and 8 with genome-wide significant enrichment of IBD sharing in atherosclerosis cases.

In a â□□gold standardâ□□ set of 112 genes with 207 established, highly penetrant pathogenic variants for Mendelian diseases, all but one were detected by applying IBDMap to the traits in BioVU. In GWAS of the traits in the same samples using Firth regression in PLINK2, only 83% of the genes had at least one variant significantly associated with the trait, demonstrating greater power for detection of these highly penetrant loci in IBD mapping. We evaluated the IBDMap haplotype filtering approach with these 112 genes, and found that it resulted in an ~2-fold increase in -log10 p-value in the same BioVU samples. Our results demonstrate the potential of our IBD mapping approach for gene discovery.

PrgmNr 3567 - Individuals with limited healthcare interactions may degrade polygenic risk score analyses

View session detail

Author Block: **S. Kulm**[1], A. R. Marderstein[1], J. G. Mezey[2], O. Elemento[1]; [1]Weill Cornell Med., New York, NY, [2]Cornell Univ, Ithaca, NY
Disclosure Block: **S. Kulm:** None.
Polygenic risk scores (PRSs) are a simple to understand, easy to produce measure of an individual's genetic susceptibility to disease. While studies have shown that individuals with high PRSs are significantly more likely to experience disease, many do not. These false positives hinder clinical integration of PRSs. As few, if any, studies have examined individuals with extreme PRSs in a systematic way, we endeavored to find features that associate to this false positive status.
In the UK Biobank, we computed PRSs for 23 diseases, selected through cross validation in a 246,062 person training phase. Within a 162,119 person testing phase the addition of a PRS to a logistic regression model containing age and sex significantly increased the corresponding AUC of 21 diseases. For each disease, we then compared 100 true positive individuals with the highest PRSs to 100 false positive individuals with the highest PRSs. Examining longitudinal features across all diseases revealed that false positives were more likely to be younger (69.4 to 71.4 to years, P -16), visit the hospital less often (7.4 to 15.2 visits over the course of the study, P -16), and visit the hospital less recently (4.8 to 3.2 years since last visit, P -16). Next, 9107 non-longitudinal, curated features were analyzed by regressing each to the disease status of the 200 individuals with extreme PRSs. Common significant features included the number of self-reported non-cancer illnesses (significant to 6 diseases), followed by self-report of other serious medical condition (5 diseases), overall health rating (3 diseases), and consumption of aspirin (3 diseases). Certain disease-specific features were noteworthy, and may indicate that some individuals may have incorrectly self-reported their disease status, including bread type for celiac disease ($P = 1.8 \times 10^{-6}$), insulin use for type II diabetes ($P = 6.9 \times 10^{-7}$) and urate levels for gout ($P = 9.7 \times 10^{-12}$).
Qualitative studies of comprehensive biographies of the highest PRS false positives illuminated a common characteristic among many significant features: frequency of interactions with the health system. This finding may be interpreted to suggest that some false positive individuals with extreme PRSs may have a disease but are not observed in a way that creates a biobank recognized diagnosis - a problem likely not unique to the UK Biobank. Accurately identifying and removing such individuals from analyses may further improve PRS research in general and expedite its uptake into clinical practice.

PrgmNr 3568 - Interactive Visualization of Relatedness and Ancestry Inference

View session detail

Author Block: **Z. Zhu**, A. W. Manichaikul, S. S. Rich, W-M. Chen; Univ. of Virginia, Charlottesville, VA
Disclosure Block: **Z. Zhu:** None.
The algorithms for estimating kinship and Identical-By-Descent (IBD) segments in the KING software tool have been widely used to identify relatedness in a genetic dataset. Various visualization plots produced from KING have been informative in presenting the inferred relatedness both at the individual level and at the summary level for the entire data, and have provided further inference and understanding of a genetic dataset beyond the kinship and IBD algorithms. However, limitations still exist, including: 1) In a summary plot for all pairs of relatedness, it is not straightforward to view the detailed IBD segments of a particular relationship pair, as well as to identify to which samples the relationship corresponds, *e.g.,* outlying relatedness is of great interest to data QC and requires follow-up; and 2) In the non-interactive plots, it is not feasible to integrate individual-level IBD segments together with the pedigree information, a feature that would be especially valuable for large pedigrees, since manual integration of various sources of information would be labor-intensive and inefficient. To address these needs, we developed four stand-alone interactive visualization applications (github.com/chenlab-uva/InteractivePlots) that facilitate the existing KING inferences using the R package â□□Shinyâ□□. A user can use the KING output files as input, interact with various summary plots by clicking a dot/line, and then display individual-level details (*e.g.*, all IBD segments of a pair of individuals) instantly on the computer screen. The four applications include: 1) viewing IBD segments interactively in a family. The documented pedigree is plotted next to the inferred relatedness graph (squares/circles for samples and lines for relatedness) which is interactive and able to return the detailed IBD segments when a relatedness line is clicked; 2) viewing IBD segments interactively in the summary IBD plot; 3) viewing the individual-level ancestry information interactively in a summary ancestry plot; and 4) viewing the individual-level Run-Of-Homozygosity (ROH) segments interactively in a summary ROH plot. We applied our interactive visualization applications to ~50,000 UK Biobank samples in families. The time to load data was only a few seconds while the interactive time was instant (

PrgmNr 3569 - Interpretable prioritization of causal genes in unsolved GWAS loci

View session detail

Author Block: **N. Cheng**[1], B. L. Trippe[1,2], C. R. Albors[1], J. C. Ulirsch[1], R. Elzur[1], E. Weeks[1], H. Finucane[1,3]; [1]Broad Inst. of MIT and Harvard, Cambridge, MA, [2]Massachusetts Inst. of Technology, Cambridge, MA, [3]Massachusetts Gen. Hosp., Boston, MA
Disclosure Block: **N. Cheng:** None.
The polygenic priority score (PoPS) is a statistical approach that leverages polygenic enrichments of gene features to prioritize causal genes at uncharacterized GWAS loci [Weeks et al. 2020 medRxiv]. However, it is still a challenge to gain biological insights using this approach. Here, we modify PoPS to use fewer features and subsequently obtain more interpretable gene prioritization models. Using summary statistics for 58 phenotypes from the UK Biobank, and for schizophrenia from the Psychiatric Genomics Consortium, we compute gene-level association scores using MAGMA [de Leeuw et al. 2015 PLOS Computational Biology]. Then we perform forward stepwise linear regression selecting 50 features to fit the MAGMA scores, producing a predicted MAGMA score for each phenotype, which we call the PoPS-FS (forward stepwise) score.
PoPS-FS performs similarly to PoPS on a gold standard gene prediction task, despite using far fewer features per phenotype (50 vs. 3,000-20,000). We obtain a set of 340 gold standard genes using fine-mapped coding variants for the 58 UK Biobank phenotypes [Ulirsch and Kanai, *In preparation*]. Prioritizing the gene with the highest PoPS-FS score in each GWS locus with a gold standard gene nearby, PoPS-FS correctly prioritizes the gold standard gene 47% of the time (PoPS accuracy = 52%). Prioritizing genes that are highest scoring and also closest to the index variant in their loci, the precision of PoPS-FS goes up to 81% (PoPS precision is also 81%). We call such genes â□□high-confidence prioritized genes.â□□
To interpret our predictions, we decompose each PoPS-FS score into a matrix of contributions from genes and features. This allows us to gain insight into our modelâ□□s decision-making process and potentially learn which genes implicate similar causal mechanisms. For instance, in schizophrenia we learn that the high score our model gives to *CACNA1C*â□□a well-known candidate causal risk geneâ□□is almost entirely due to a feature which annotates the protein-protein interactions with *CACNA2D1.*
In HbA1c, we find that *SMIM1*, *ANK1*, and *SPTB*â□□three high-confidence prioritized genes related to red blood cell formation and functionâ□□are among each otherâ□□s top 10 contributing genes. Furthermore, *G6PC2*, *GCKR*, and *TCF7L2*â□□three high-confidence prioritized genes related to glucose metabolism and type 2 diabetesâ□□are among each otherâ□□s top 10 contributing genes. This suggests the presence of at least two distinct pathwaysâ□□one erythrocytic and the other glucose-relatedâ□□that regulate HbA1c levels, consistent with previous genetic studies of this phenotype. Designing procedures to validate these decompositions is a key future direction for this work.

PrgmNr 3570 - LD Score regression (LDSC) with more comprehensive catalogs of LD scores

View session detail

Author Block: **W. Liu**, B. Rowland, J. D. Rosen, J. Wen, Y. Li; Univ. of North Carolina at Chapel Hill, Chapel Hill, NC
Disclosure Block: **W. Liu:** None.
Background: LD Score regression (LDSC) is widely used for estimating the degree of inflation in genome-wide association studies (GWAS) test statistics, by distinguishing between inflation from polygenicity and confounding biases using GWAS summary statistics. In addition, LDSC estimates heritability, genetic correlation, and partitioned heritability for prioritizing functional annotations.
Objectives: The authors of LDSC released LD scores for ~1.3 million HapMap3 variants separately for European and East Asian samples from the 1000 Genomes Project (1000G) covering only a subset of common variants and few low frequency or rare variants (LFRV). We investigated how LDSC estimates change when including additional variants and using a more comprehensive LD reference panel built from whole genome sequencing data from the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program.
Methods: We computed LD scores using LD calculated for TOPMed European and East Asian subjects and collected GWAS summary statistics for 40 human diseases and traits including BMI, height, blood cell traits, and brain-related disorders and traits for Europeans, and 13 blood cell traits for East Asians. We compared 1000G LD scores and TOPMed LD scores for HapMap3 variants to assess the impact of including additional LFRV in LD scores on LDSC estimates. Next, we compared LDSC estimates using TOPMed LD scores from variants with MAF>1% in TOPMed to only HapMap3 variants. Finally, we compared LDSC estimates using TOPMed LD scores from LFRV to common variants.
Results: We found 10 and 8 studies exhibited evidence of inflation due to confounding bias (LDSC intercept > 1.1, genomic control inflation factor $\hat{\lambda}_{GC}$ > 1.1) using TOPMed LD scores, but not with 1000G LD scores, for European and East Asian GWAS summary statistics, respectively. The LDSC intercept estimate using TOPMed LD scores was 3.1%-39.9% larger than estimates from the 1000G LD scores. In addition, using additional variants with MAF>1% in TOPMed decreased the estimate of the LDSC intercept (0.2%-9.8%). Little inflation for LFRV was observed.
Conclusions: Our evaluations comprehensively assessed the impact of including more LFRV in LD score computation on quantification of inflation factor and will guide the choices of the most appropriate LD scores for different ancestry and/or in different allele frequency categories.

PrgmNr 3571 - Leveraging functional annotation in a fast and powerful eQTL weighted gene-based association test

View session detail

Author Block: **T. Haynes**[1], L. Tran[2], S. Gonzales[2], X. Wang[3]; [1]Texas Academy of Math & Sci., Denton, TX, [2]Univ. of North Texas, Denton, TX, [3]Univ North Texas, Denton, TX
Disclosure Block: **T. Haynes:** None.
With the success of genome-wide association study (GWAS) and next generation sequencing data analysis, vast amounts of GWAS summary data are publicly available. To further utilize the GWAS summary data and increase statistical power in genetic association studies, people have proposed a number of transcriptome-wide association study (TWAS) type methods which use gene expression as a mediating trait linking genetic variations with diseases. Usually, these methods employ two key steps: 1) predicting gene expression levels based on inferred gene expression quantitative trait loci (eQTLs); 2) identifying expression-mediated genetic effects on diseases by associating phenotypes with predicted gene expression levels. The success of these methods critically depends on the first step - identification of eQTLs. However, some eQTLs may not be functional in the corresponding tissue, due to linkage disequilibrium (LD) and the correlation of gene expression between tissues. To increase the interpretability and meaningfulness of the genes detected, we propose an omnibus Test to identify disease-associated Genes leveraging Epigenetic information (TGE). TGE utilizes a Cauchy association test to integrate association evidence demonstrated by three different traditional tests (burden test, quadratic test, and adaptive test) using GWAS summary data with multiple eQTLâ□□derived weights. Through prioritizing genetic variants with tissue-specific epigenetic annotation, TGE can better identify genetic variants that are statistically predictive and biologically functional. The p value of the proposed test can be calculated analytically, and thus it is fast and efficient. We applied our proposed test to two schizophrenia (SCZ) GWAS summary data sets and two lipids trait (HDL) GWAS summary data sets. A significantly higher percentage of eQTLs identified by TGE are inferred to be functional and more are deleterious based on their Combined Annotation Dependent Depletion (CADD) scores. Compared with existing methods, our proposed TGE can identify more traitâ□□associated meaningful genes.

PrgmNr 3572 - Leveraging Health Systems Data to Characterize a Large Effect Variant Conferring Risk for Liver Disease in Puerto Ricans

View session detail

Author Block: **G. M. Belbin**[1,2], S. Rutledge[1,2], T. Dodatko[1], S. Cullina[1,2], M. C. Turchin[1,2], S. Kohli[1,2], D. Torre[1], M-C. Yee[3], C. R. Gignoux[4], N. S. Abul-Husn[1,2], S. Houten[1], E. E. Kenny[1,2]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]The Inst. for Genomic Hlth., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Stanford Univ., San Francisco, CA, [4]Univ. of Colorado, Denver, CO
Disclosure Block: **G.M. Belbin:** None.
Broad-scale adoption of genomic data in health systems offers opportunities for leveraging population genetics to better understand disease risk. We explored patterns of Identity-by-Descent (IBD) sharing in a patient population in New York City and noted elevated levels of sharing in Puerto Rican (PR) ancestry participants. In an extension of population-based linkage mapping, we clustered IBD haplotypes by homology within the PR group and used these clusters to perform a phenome-wide association study, where we systematically explored relationships between shared IBD haplotypes and over 10,000 health outcomes derived from the electronic health records. In doing so we uncovered one association that achieved study-wide significance between an IBD-haplotype spanning the ABCB4 locus and severe liver disease. We used genome sequencing and in silico approaches to fine-map the signal to a non-coding variant (c.2784-12T>C) in the gene ABCB4. In vitro analysis confirmed the variant disrupted splicing of the ABCB4 pre-mRNA. Four of five homozygotes had evidence of advanced liver disease, and there was a significant association with liver disease among heterozygotes. Examination of local ancestry on the background of the significant IBD haplotype determined the haplotype to be of European origin. Population-level screening revealed the variant to be at a carrier rate of 1.95% in PR individuals, and extremely rare within some European populations, while otherwise absent globally, suggestive of a PR founder effect. This work demonstrates that integrating EHR and genomic data at a population-scale can facilitate novel strategies for understanding the continuum of genomic risk for common diseases, particularly in populations underrepresented in genomic medicine.

PrgmNr 3573 - Leveraging simulations in extended pedigrees for population genetic analysis

View session detail

Author Block: **L. Anderson-Trocme**[1], S. Zabad[2], W. Wohns[3], I. Krukov[4], A. Diaz-Papkovich[2], C. Dina[5], S. Gravel[6]; [1]McGIll Univ., Montreal, QC, Canada, [2]McGill Univ., Montreal, QC, Canada, [3]Broad Inst., Cambridge, MA, [4]Montreal, QC, Canada, [5]Nantes, France, [6]Mc Gill Univ, Montreal, QC, Canada
Disclosure Block: **L. Anderson-Trocme:** None.
Family scale pedigrees have provided useful frameworks for understanding the transmission histories of mendelian diseases. At a population scale, they are increasingly used to bridge the gap between individuals and populations. The present study considers the extended pedigree of the French Canadian population of Quebec compiled from over six million parish records dating back to the first French settlers four centuries ago. We use an extension of msprime [1] to generate genealogically aware simulations at a population scale and evaluate their accuracy by comparing them to over seven thousand genotyped individuals linked to the pedigree. Simulations within the genealogy capture the extensive population structure in the populations, bridging the gap between family and population studies. We show excellent agreement with empirical measures of diversity (such as PCA) and contrast inferred tree sequences from both real and simulated data using a site specific genealogical nearest neighbour statistic [2,3]. Using the spatial information in the pedigree, we show how anisotropic range dispersal and serial founder events have influenced present day population structure. Using the concordance between haplotype-based nearest neighbours with pedigree-based kinship, we explore novel frameworks to study realistic population dynamics and the transmission of risk variants, as well as the inference of ancestral haplotypes.
References : **[1]** Nelson, et al. PLoS genetics (2020). **[2]** Kelleher, et al. Nature genetics (2019). **[3]** Wohns, et al. bioRxiv (2021).

PrgmNr 3574 - MAST-Decon: Smooth Cell-type Deconvolution Method for Spatial Transcriptomics Data

View session detail

Author Block: **T. Luo**[1], J. Chen[1], H. Zhu[1,2], Y. Li[1,3,4]; [1]Dept. of Biostatistics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [2]BioMed. Res. Imaging Ctr., Sch. of Med., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [3]Dept. of Genetics, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [4]Dept. of Computer Sci., Univ. of North Carolina at Chapel Hill, Chapel Hill, NC
Disclosure Block: **T. Luo:** None.
Spatial transcriptomics (ST) technologies have gained increasing popularity due to their ability to provide positional context of gene expressions in a tissue. One major limitation of current commercially available ST methods such as the 10X Genomics Visium platform is that they cannot yet reach single cell resolution. The number of cells within a spatial spot may range from 1 to 200 depending on the biological tissue and ST platform. Therefore, any downstream analysis such as spatially variable gene detection could be confounded by differential cell type compositions across spots. Cell type deconvolution for ST data is critical in order to fully reveal underlying biological mechanisms. Existing ST data deconvolution methods share two common limitations: first, few of them utilize spatial neighborhood information. Existing methods such as RCTD and SPOTlight intrinsically treat each spatial spot as independent of neighboring spots, although we anticipate nearby spots to share similar cell type compositions based on clinical evidence of tissue structures. Such limitation could be amplified when sequencing depths at single spots are relatively low so that borrowing information from neighboring spots is necessary in order to obtain reliable deconvolution results. Second, although Visium data provide us with a histological image which could add additional information regarding spot heterogeneity, most existing methods do not utilize this H&E image. To solve these two limitations, we developed Multiscale Adaptive ST Deconvolution (MAST-Decon), a smooth deconvolution method for ST data. MAST-Decon uses a weighted likelihood approach and incorporates both gene expression data, spatial neighborhood information and H&E image features by constructing different kernel functions to obtain a smooth deconvolution result. We showcased the strength of MAST-Decon through simulations based on real data, including a single-cell dataset of mouse brain primary visual cortex. By introducing spatial smoothness, we were able to correct several spots erroneously inferred by RCTD when the average UMI count per spot is at ~2.5k level. Overall, we were able to improve the Spearman correlation between deconvolution results and ground truth cell-type proportions from 0.757 (RCTD) to 0.823 (MAST-Decon), and reduce RMSE between deconvolution results and ground truths from 0.0491 (RCTD) to 0.0334 (MAST-Decon).

PrgmNr 3575 - Methods for genetic association analysis of mother-child pairs

View session detail

Author Block: **D. Ray**[1], C. Vergara[1], M. Taub[2], G. L. Wojcik[1], C. M. Ladd-Acosta[3], T. H. Beaty[3], P. Duggal[3]; [1]Epidemiology, Johns Hopkins Univ., Baltimore, MD, [2]Johns Hopkins, Baltimore, MD, [3]Johns Hopkins Univ., Baltimore, MD
Disclosure Block: **D. Ray:** None.
Genetic association studies of pediatric outcomes often employ family-based designs. One of the most popular family-based designs is the case-parent trio design that considers a nuclear family of parents and their affected child. This trio design is particularly advantageous for studying relatively rare disorders because amassing adequate numbers of affected children require sampling from multiple populations. This can cause type I error inflation due to population stratification if a population-based design is used. However, genetic information on both parents is often not available; fathers may be missing as they are harder to recruit, and paternity is more difficult to ensure than maternity. While there are a few statistical methods for analyzing mother-child dyad data, it is not clear if one method is advantageous over another under different dyad designs (e.g. case-mother dyads, case-mother/control-mother dyads), and if they provide the same advantage as methods for trio design in protecting against population stratification. Further, type I error and statistical power of most of these methods were evaluated at nominal significance levels only. Recent benchmarking articles on family-based genetic studies have either focused on trios and larger pedigrees or considered only one class of methods. Here we reviewed existing methods for analyzing genome-wide data on dyads and performed extensive simulation experiments to benchmark their type I error and power at stringent error levels against the corresponding trio design under different scenarios. We considered samples from homogenous racial/ethnic populations as well as assumed scenarios where population substructure exists. We also reviewed and evaluated methods for analyzing case-parent trios and mother-child dyads together. We applied these methods on genotyped/imputed data from multi-ethnic mother-child pairs only or trios only (ascertained through a child affected by nonsyndromic cleft lip with or without cleft palate) or a combination of both dyads and trios from the Gene, Environment Association Studies consortium (GENEVA). Results from the GENEVA study corroborate findings from our simulation experiments. Finally, we provide recommendations on which statistical method for dyads to use and under what scenario. This work is motivated by the Environmental influences on Child Health Outcomes (ECHO) study where genome-wide data, mostly on mother-child pairs, are becoming available. Our recommendations will not only help ECHO to select the protocols to use for discovering genetic risk factors of pediatric outcomes but also provide guidelines for other studies with dyad designs.

PrgmNr 3576 - Multiethnic joint analysis of summary statistics from genome-wide association studies

View session detail

Author Block: **J. Shen**[1], L. Jiang[1], K. Wang[1], P. J. Newcombe[2], C. Haiman[1], D. V. Conti[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]MRC Biostatistics Unit, Cambridge, United Kingdom
Disclosure Block: **J. Shen:** None.
Over the last 20 years, Genome-wide Association Studies (GWAS) have been able to identify genetic regions associated with traits and diseases in different populations. As a post-GWAS approach, multiethnic fine-mapping aims to identify underlying causal variants by combining summary statistics and leveraging different linkage disequilibrium (LD) structures across diverse populations. Here, we expand upon our previous approach for single-population fine-mapping through Joint Analysis of Marginal SNP Effects (JAM) to a multiethnic analysis (mJAM). Under the assumption that true causal variants are common across populations, our joint model explicitly incorporates the different LD structures across populations and yields a conditional fixed-effect meta-analysis (FE). To pinpoint causal variants from highly correlated signals efficiently, we incorporate Sum of Single Effects (SuSiE), a Bayesian stepwise selection method, within the mJAM framework. Through simulation studies based on realistic effect sizes and levels of LD, we demonstrate that mJAM performs better than other existing multi-ethnic methods including FE, conditional and joint analysis using summary data (COJO), and multiple study causal variants identification in associated regions (MsCAVIAR). The flexible mJAM framework can be extended to deal with binary disease outcome or missing SNP information in some populations, which is a unique advantage of mJAM over other existing methods. In a real data application, we apply mJAM to recently published summary statistics from a trans-ancestry prostate cancer GWAS.

PrgmNr 3577 - Multi-trait analysis of gene-environment-wide association using MAGENTA

View session detail

Author Block: **L. Luo**[1], J. Shen[1], Z. Tang[2]; [1]Merck & Co., Inc., Rahway, NJ, [2]Univ. of Wisconsin-Madison, Madison, WI
Disclosure Block: **L. Luo:** None.
Many complex traits (i.e., diseases or drug responses) are influenced by the interplay of genetic and environmental factors. Detecting genotype-by-environment interaction (GEI) is fundamental in understanding their etiology and increasing the power to detect complex-trait associated genes. However, it is challenging to identify GEI signals largely due to small effect sizes and limited sample sizes. Aggregating association evidence across multiple variants and multiple traits helps boost the power. Here, we propose MAGENTA (**M**ulti-trait **A**nalysis of **G**ene-**EN**vironmen**T**-wide **A**ssociation), a novel random-effects meta-analysis framework for testing GEI effect and the genetic main and GEI joint effects respectively across a SNP-set and multiple traits. MAGENTA is motivated by the observation that tests for the GEI are closely related to the test of heterogeneity in random-effects meta-analysis with strong heterogeneous genetic effects among environmental groups implying the existence of GEI effects. Compared with single-trait based approaches, MAGENTA gains more power by using genetic correlation information to accommodate the heterogeneous genetic effects among traits. By using summary statistics data across environmental groups as input, MAGENTA is computationally fast and efficient while analyzing large-scale genome-wide GEI association studies. Our extensive simulation studies show MAGENTA controls type I error well and yields stable power across various scenarios. We applied MAGENTA to the UK-Biobank whole exome sequencing data (N=46376) and identified a significant GEI effect between APOE gene and sex on three lipid traits (HDL, LDL and triglycerides), which was missed by single-variant GEI test or other single-trait SNP-set based methods (i.e., MAGEE).

PrgmNr 3578 - Multivariate modeling of direct and proxy GWAS indicates substantial common variant heritability of Alzheimerâ□□s disease

View session detail

Author Block: **J. de la Fuente**[1,2], A. D. Grotzinger[1,3], R. E. Marioni[4], M. G. Nivard[5], E. M. Tucker-Drob[1,2]; [1]Dept. of Psychology, Univ. of Texas at Austin, Austin, TX, [2]Population Res. Ctr. and Ctr. on Aging and Population Sci., Univ. of Texas at Austin, Austin, TX, [3]Psychiatric and Neurodevelopmental Genetics Unit (PNGU) and the Ctr. for Genomic Med., Massachusetts Gen. Hosp., Boston, MA, [4]Ctr. for Genomic and Experimental Med., Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom, [5]Dept. of Biological Psychology, VU Univ. Amsterdam, Amsterdam, Noord Holland, Netherlands

Disclosure Block: **J. de la Fuente:** None.

Genome-wide association studies (GWAS) of proxy-phenotypes using family history of disease (GWAX) can substantially boost power when combined with traditional case-control GWAS across a range of disease traits. The usefulness of GWAX for enhancing GWAS discovery has been particularly noticeable in the context of late-onset Alzheimerâ□□s Disease (AD). However, there is a stark contrast between heritability estimates derived from twin-based approaches and combined GWAX-GWAS meta-analysis of AD. While heritability estimates from twin approaches are approximately 60%, combined GWAX-GWAS meta-analysis of AD indicate very low common variant SNP heritability estimates after excluding the *APOE* region, with the most recently reported estimate of 2.5% on the liability scale coming from the application of LD Score Regression (LDSC) to the largest sample to date. Here, we demonstrate that commonly used approaches for combining GWAX and GWAS data produce dramatic underestimates of heritability, and we introduce a multivariate method for estimating individual SNP effects and recovering unbiased estimates of SNP heritability when combining GWAS and GWAX summary data. Via simulation, we show that our method recovers unbiased estimates of common variant SNP heritability and of individual SNP effects under a variety of circumstances in which standard approaches, including those that correct for the indirect nature of the GWAX, dramatically underestimate these quantities. Using recently released GWAS and GWAX summary data for Alzheimerâ□□s disease (AD), our method estimates the common variant SNP heritability of AD excluding the *APOE* region to be more than double (~6-10%) recent estimates using the same assumptions regarding population prevalence. Moreover, we show the corresponding estimate for biological AD (based on prevalence rate estimates from recently published molecular imaging data) as high as ~20%. Finally, our local SNP heritability analyses indicate that more than 70% of common variant SNP heritability of AD exists outside the *APOE* region, evidencing a polygenic signal that is relatively diffusely distributed across the genome.

PrgmNr 3579 - Novel Global Quantile Regression Based Method for Differential Gene Expression Analysis of Quantitative traits

View session detail

Author Block: **S. Tang**[1], A. S. Buchman[2], P. L. De Jager[3], D. A. Bennett[4], Q. Zheng[5], J. Yang[6]; [1]Emory Univ., Atlanta, GA, [2]Rush Alzheimerâ□□s Disease Ctr., Chicago, IL, [3]Columbia Univ Med Ctr, New York, NY, [4]Rush Univ. Med. Ctr., Chicago, IL, [5]Univ. of Louisville, Louisville, KY, [6]Emory Univ, Atlanta, GA
Disclosure Block: **S. Tang:** None.
High-throughput RNA sequencing (RNA-seq) has been widely used in studies of complex diseases to detect differentially expressed genes with respect to complex traits and diseases. Existing popular methods such as DESeq2 and edgeR are only developed for studying dichotomous traits, which are based on generalized linear regression models with RNAseq read counts as the response variable. Recently, a standard linear regression based test method was proposed for differential gene expression (DGE) analysis of quantitative traits, which takes the trait of interest as the response variable and the RNA-seq read counts as the test covariate.
However, standard linear regression model is limited to only test the association with respect to the mean of the response variable and the test covariate. To account for heterogeneous associations with respect to different quantiles through the entire distribution of the trait, we propose a novel global quantile regression based method to test for differential gene expressions with quantitative traits (GQR-DEseq). GQR-DEseq uses the quantile rank-score test that is an analogous variance component score test to account for response-covariate association across the whole quantile domain of the trait. By simulation studies using real RNAseq data, we compared the performance of GQR-DEseq with the DGE method based on standard linear regression. We considered a monotone negative increasing and a monotone positive decreasing effect size functions (functions of quantiles) estimated from our real data as well as a zero effect size function for null models. Regardless of the direction of effect size functions, GQR-DEseq had higher power than standard linear regression based tests, 54.1% vs. 2.7% with positive effect size, 99.6% vs. 87.8% with negative effect size. Our null simulations showed GQR-DEseq controlled for type I error. These simulation studies show that GQR-DEseq has higher power than the standard linear regression based tests when the covariate-response effect size is heterogeneous with respect to response quantiles.
Further, we applied GQR-DEseq to the RNA-seq data of frontal cortex tissues for studying Alzheimerâ□□s dementia (AD), taking the quantitative rate of cognitive decline as the trait of interest. GQR-DEseq detected 47 significant genes with FDRADAMTS-2 (p-value =3.24Ã□10^{-15}) has been suggested to be a therapeutic target for adult brain disorders such as schizophrenia and Alzheimer's disease.
A free tool of GQR-DEseq will be available on Github which provides a novel resource to detect DGE of quantitative traits.

PrgmNr 3581 - Reconstructing SNP Allele and Genotype Frequencies from GWAS Summary Statistics

View session detail

Author Block: **Z. Yang**, P. Paschou, P. Drineas; Purdue Univ., West Lafayette, IN
Disclosure Block: **Z. Yang:** None.

The emergence of genomewide association studies (GWAS) has led to the creation of large repositories of human genetic variation, creating enormous opportunities for genetic research and worldwide collaboration. Methods that are based on GWAS summary statistics seek to leverage such records, overcoming barriers that often exist in individual-level data access while also offering significant computational savings. Here, we propose a novel framework that can reconstruct allelic and genotypic counts/frequencies for each SNP from case-control GWAS summary statistics. Our framework is simple and efficient without the need of any complicated underlying assumptions. Illustrating the great potential of this framework we also propose three summary-statistics-based applications implemented in a new software package (ReACt): GWAS metaanalysis (with and without sample overlap), case-case GWAS, and, for the first time, group polygenic risk score (PRS) estimation. We evaluate our methods against the current state-of-the-art on both synthetic data and real genotype data and show high performance in power and error control. Our novel group PRS method based on summary statistics could not be achieved prior to our proposed framework. We demonstrate here the potential applications and advantages of this approach. Our work further highlights the great potential of summary-statistics-based methodologies towards elucidating the genetic background of complex disease and opens up new avenues for research.

PrgmNr 3582 - SparsePro: an ultra-efficient Bayesian fine-mapping method integrating summary statistics and functional annotations

View session detail

Author Block: **W. Zhang**, Y. Li; McGill Univ., Montreal, QC, Canada
Disclosure Block: **W. Zhang:** None.

Identifying causal variants from genome-wide association studies (GWASs) is challenging due to widespread linkage disequilibrium (LD). Functional annotations of the genome may help prioritize variants that are biologically relevant, thus improve fine-mapping of GWAS results. However, canonical fine-mapping methods incorporating functional annotations have a high computational cost, particularly when the underlying genetic architecture and LD pattern are complex.

We propose SparsePro, to efficiently conduct functionally informed statistical fine-mapping. Our method enjoys three innovations: First, by creating a sparse low-dimensional projection of the high-dimensional genotype space, we enable a linear direct search of causal variants instead of an exponential search of causal configurations used in existing methods; Second, we adopt a probabilistic framework with an efficient variational expectation-maximization algorithm to integrate statistical associations and functional annotations; Last, we specifically test for the relevance of functional annotations, which improve the robustness when trait-related functional categories are unknown.

We evaluated SparsePro through extensive simulations using UK Biobank genotypes. Under various settings of causal configuration, heritability, LD pattern, and intensity of functional annotation enrichment, SparsePro achieved superior or comparable discriminative power in identifying causal variants compared to state-of-the-art methods, with much improved computational efficiency. With a five-fold enrichment of functional annotations on causal variants, SparsePro achieved the highest overall area under the precision-recall curve (AUPRC) of **0.786**, outperforming the runner-up method PAINTOR with an AUPRC of 0.654. Notably, SparsePro was also **455** times faster than PAINTOR. With irrelevant functional annotations, SparsePro achieved an AUPRC of 0.604. While the runner-up methods, FINEMAP and SuSIE, achieved the same AUPRC, SparsePro was 15 and 44 times faster, respectively. We applied SparsePro to conduct functionally informed fine-mapping of complex trait GWASs, including body mass index, height, type 2 diabetes and breast cancer. We have discovered relevant functional enrichment and promising candidates of causal variants.

In summary, we have developed an ultra-efficient and accurate Bayesian fine-mapping method with the ability to integrate summary statistics and functional annotations. Our method will have broad utility in dissecting disease-relevant functional elements, identifying target genes and increasing the yield of functional follow-up of GWASs.

PrgmNr 3583 - The power of pathway-specific polygenic risk scores

View session detail

Author Block: **S. Choi**[1], J. Garcia-GonzÃ¡lez[1], Y. Ruan[2], J. Johnson[1], H. Wu[1], O. Paul[1]; [1]Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Broad Inst., Cambridge, MA
Disclosure Block: **S. Choi:** None.
Polygenic risk scores (PRSs) have been among the leading advances in biomedicine in recent years. As a proxy of genetic liability, PRSs are utilised across multiple fields and applications. While numerous statistical and machine learning methods have been developed to optimize their predictive accuracy, all of these distil genetic liability to a single number based on an aggregation of an individualâ□□s genome-wide alleles. This results in a key loss of information about an individualâ□□s genetic profile, which could be critical given the functional sub-structure of the genome and the heterogeneity of complex disease. Here we evaluate the performance of pathway-specific PRSs, in which *k* polygenic scores are calculated across *k* genomic pathways for each individual, and introduce a software, *PRSet*, for computing and analysing pathway PRS. We find that pathway PRSs have similar power for evaluating pathway enrichment of GWAS signal as the leading methods, with the distinct advantage of providing estimates of pathway genetic liability at the individual-level. Exemplifying their utility, we demonstrate that pathway PRSs can stratify diseases into sub-types in the UK Biobank with substantially greater power than genome-wide PRSs. Compared to genome-wide PRSs, we expect pathway-specific PRSs to offer greater insights into the heterogeneity of complex disease and treatment response, generate more biologically tractable therapeutic targets, and provide a more powerful path to precision medicine.

PrgmNr 3585 - Using metagenomics to assess global levels of antimicrobial resistance

View session detail

Author Block: **S. T. Eckert**[1], X. Zhan[2], D. J. Liu[1]; [1]Penn State Coll. of Med., Hershey, PA, [2]UT Southwestern, Dallas, TX
Disclosure Block: **S.T. Eckert:** None.
Antimicrobial resistance (AMR) poses a major public health challenge that leads to nearly 3 million infections and over 35,000 deaths per year in the US alone. This estimate considers only symptomatic instances of antimicrobial-resistant microbes, but as we have learned from recent developments in infectious disease, latent, asymptomatic carriers of infection are much more common than previously thought. These unaffected carriers likely shape the landscape of how AMR infections spread. It is important to gain a deeper understanding of asymptomatic AMR in the general population. Numerous publicly available metagenomic sequencing datasets offer a great opportunity to quantify the extent of asymptomatic AMR infection in the general population. We previously developed a database, VAMPr, which uses pairs of bacterial sequences and antibiotic-resistant/susceptible status as training data. A machine learning model was obtained that accurately predicts AMR status from bacterial sequence variants. To apply VAMPr to metagenomic sequences, we further developed a pipeline, MetaPrism2, to call microbial, amino acid variants from metagenomic sequencing datasets. The pipeline first assembles contigs from metagenomic sequences, which are then realigned to the contigs of known origin. A consensus sequence is then generated from the called variants and contigs, which is then fed into VAMPr to predict AMR status for each bacteria-antibiotic pair. To apply this pipeline, we called amino acids variants and predicted AMR status for a dozen curated metagenomic datasets, totaling 2532 individuals and 10016 samples from various countries. An estimated 94.8% of individuals latently carry at least one bacterial species that is resistant to at least one antibiotic. We further used logistic regression to associate predicted AMR status with various covariates from meta-information including (but not limited to) antibiotic use, age, sex, BMI, and the study condition. Direction of effect for association with antibiotic use depended on which antibiotics the subjects were prescribed. For example, penicillin-class antibiotic use generated more antibiotic resistance to tobramycin (OR=1.55, 95% CI[1.01,2.38], p=0.04), while usage of other antimicrobial groups may lead to increased susceptibility to tobramycin (OR=0.56, 95% CI[0.33,0.97], p=0.037). Another significant association of interest is that of antibiotic resistance with a modifiable risk factor, smoking (clindamycin, OR=5.26, 95% CI [1.15,24.1], p=0.03). Our methods, results, and continued investigation will elucidate the global distribution for AMR status and pinpoint the risk factors for AMR infection.

PrgmNr 3586 - Utility of admixed individuals in genome wide association studies: A test case in COVID-19

View session detail

Author Block: **S. Tuminello**[1,2], N. Hu[1], H. Berk-Rauch[1], P. Cotiza[3], L. Boytard[3], C. Arguelles-Grande[3], P. Zappile[4], E. Guzman[4], A. Heguy[4], I. Osman[3], L. Thorpe[2], A. Chakravarti[1]; [1]Ctr. for Human Genetics and Genomics, New York Univ. Sch. of Med., New York City, NY, [2]Div. of Epidemiology, Dept. of Population Hlth., New York Univ. Sch. of Med., New York City, NY, [3]The New York Univ. Langone Hlth.(NYULH) Ctr. of Biospecimen Res. and Dev., Office of Sci. and Res., New York Univ. Sch. of Med., New York City, NY, [4]Genome Technology Ctr., Div. of Advanced Res. Technologies, New York Univ. Sch. of Med., New York City, NY

Disclosure Block: **S. Tuminello:** None.

Human genetic studies usually stratify subjects by continental ancestry and have historically excluded admixed individuals; yet more than a third of the US population is admixed and will be increasingly so. During the COVID-19 pandemic, biospecimens linked to clinical data were rapidly collected at New York University Langone Health (NYULH) irrespective of ancestry. We demonstrate here how genetic studies of cosmopolitan, admixed individuals can be informative for ancestry-specific risks of COVID-19 severity. Patients were genotyped using 1X whole genome sequencing followed by genotype imputation. 1,538 COVID-19 positive, hospitalized patients were studied in which we identified 59,862,263 genomic variants. 46% of patients were admixed, 198 (13%) were African American, 81 (5%) Hispanic, and 426 (28%) admixed non-African American or non-Hispanic. The remaining 54% were non-admixed and classified as belonging solely to one of five groups: African (193, 13%), American (10, African:0.48, $OR_{African\text{-}American}$:0.65, OR_{Asian}:1.51; rs73064425: $OR_{African}$:0.45, $OR_{African\text{-}American}$:0.63, OR_{Asian}:2.42; estimates not statistically significant). This implies that variants associated with genetic susceptibility to COVID-19 vary by genomic ancestral populations, impacting heterogeneity in outcomes between admixed individuals. We are now estimating local ancestry across the genome to conduct genetic association studies of COVID-19 to estimate ancestry-specific allelic effects.

PrgmNr 3587 - Utilizing phenotype risk scores and genetically-regulated expression in a clinical biobank to investigate the genetic architecture of Mendelian disease genes

View session detail

Author Block: **T. W. Miller-Fleming**[1], A. B. Faucon[2], L. Bastarache[3], N. J. Cox[1]; [1]Vanderbilt Univ Med Ctr., Nashville, TN, [2]Vanderbilt, Nashville, TN, [3]Nashville, TN
Disclosure Block: **T.W. Miller-Fleming:** None.
Background: Mendelian diseases typically arise from rare, coding mutations in single genes and are evaluated in small, family-based studies. Despite the estimation that more than 7,000 Mendelian conditions exist, many are poorly understood and have no treatment. We utilized large-scale data and tools to better understand the genes responsible for Mendelian disease. In this study, we leveraged the dense phenotype information in the Vanderbilt electronic health records (EHR) and biobank to identify individuals that share core features of Mendelian diseases and determine whether the genetically-regulated expression (GREX) of the causal genes were correlated with the disease features. **Methods:** We generated phenotype risk scores (PheRS) for 80,775 patients in the biobank using Human Phenotype Ontology/HPO terms in OMIM (Online Mendelian Inheritance in Man) mapped to phenotypes within the EHR for Mendelian diseases. Using linear regression models, we tested for associations between the PheRS for 2,817 distinct Mendelian diseases with the GREX of the causal gene for each disorder. **Results:** Across the 3,114 gene-disease pairs tested, we identified 292 significant associations between GREX and PheRS, suggesting that common, regulatory genetic variation can contribute to similar phenotypic consequences as rare, coding variation. Our significant findings include: Cystic Fibrosis with *CFTR* ($p=2.99x10^{-6}$), Alzheimer disease with *APOE* ($p=1.45x10^{-4}$), and Hypercholesterolemia with *APOB* ($p=2.60x10^{-4}$). The significant gene-disease pairs were not enriched for recessive or dominant inheritance patterns of disease, loss-of-function or gain-of-function genetic mechanism, or loss-of-function intolerant genes. After conditioning a subset of our analyses on diagnosis status in the EHR, we found that the significant associations were not driven by diagnosed individuals, rather that many individuals with the highest PheRS for Mendelian conditions were not diagnosed within the biobank. **Discussion:** Our findings support the utilization of clinical biobanks and gene expression imputation tools in the study of rare disease. We find that for ~10% of the Mendelian disease genes tested, common, regulatory variation contributes to phenotypic consequences consistent with our previous knowledge of the disease. Additionally, we find that many individuals within the biobank have high PheRS for Mendelian disorders, despite many being undiagnosed. This work suggests that a better understanding of Mendelian disease genes could benefit a much larger proportion of the population than those currently diagnosed with rare, monogenic diseases.

PrgmNr 3588 - A phenome-wide Mendelian randomization study of plasma triglycerides and more than 6,000 diseases and traits

View session detail

Author Block: **S. Bafna**[1,2], G. Rocheleau[1,2], I. Forrest[1,2], A. Duffy[1,2], C. Marquez-Luna[1,2], B. Petrazzini[1,2], H. Vy[1,2], D. Jordan[1,2], M. Verbanck[3], R. Do[1,2]; [1]Charles Bronfman Inst. for Personalized Med., Icahn Sch. of Med. at Mount Sinai, New York, NY, [2]Dept. of Genetics and Genomic Sci., Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]UniversitÃ© de Paris, Paris, France
Disclosure Block: **S. Bafna:** None.
Mendelian randomization studies have shown that decreased plasma triglycerides is causally related to decreased coronary artery disease (CAD) risk. This is further supported by triglyceride-lowering therapeutics which have also demonstrated protection against CAD. Phenome-wide Mendelian randomization (MR) can be used to identify additional outcomes causal to plasma triglycerides. The objective of this study is to evaluate potential causal effects of triglycerides on a wide array of disease outcomes and traits by performing a phenome-wide MR analysis in the electronic health record (EHR)-linked UK Biobank and FinnGen Biobank.

We performed phenome-wide MR analyses to test the causal effect of triglycerides on 4,156 and 2,264 diseases and traits using genome-wide association summary statistics in the Pan-UK Biobank and FinnGen Biobank. We identified 141 independent triglyceride-altering genetic variants serving as instrumental variables for four different pleiotropy-aware MR methods: inverse-variance weighted (IVW), MR-Egger and Weighted Median and MR-PRESSO. To determine statistical significance, we set *P* â□□5 for the IVW test after accounting for multiple testing. From this analysis, we discovered causality of plasma triglycerides with 223 traits and diseases in total (183 in Pan-UK Biobank and 40 in FinnGen). Of these, 161 out of the 223 exhibited evidence of horizontal pleiotropic bias (P -4 (0.05/223). We identified known causal relationships, including plasma triglycerides with ischemic heart disease (IVW Î² = 0.30, $P = 1.1Ã□10^{-19}$) and myocardial infarction (IVW Î² = 0.36, $P = 2.7Ã□10^{-15}$). We also identified novel causal relationships including plasma triglycerides with peptic ulcer (IVW Î² = 0.92, $P = 3.1Ã□10^{-8}$), cerebral aneurysm (IVW Î² = 0.93, $P = 5.7Ã□10^{-6}$), enthesopathies (IVW Î² = -0.47, $P = 1.5Ã□10^{-6}$), vertical strabismus (IVW Î² = -0.89, $P = 2.4Ã□10^{-6}$), wet age-related macular degeneration (IVW Î² = -0.58, $P = 7.1Ã□10^{-6}$), among many others.

Phenome-wide MR analyses demonstrate that plasma triglycerides is causally related to a number of diseases. Triglyceride-lowering therapeutics should be considered in the context of these additional causal conditions.

PrgmNr 3589 - Assessing the difference in multiple sclerosis age of onset amongst different racial populations through the use of electronic health records

View session detail

Author Block: **A. Hernandez**[1], F. B. Briggs[2], M. F. Davis[3]; [1]Brigham Young Univ., Provo, UT, [2]Case Western Reserve Univ. Sch. of Med., Cleveland, OH, [3]Brigham Young Univ, Provo, UT
Disclosure Block: **A. Hernandez:** None.
Multiple Sclerosis (MS) is a complex autoimmune disease where the myelin sheath that covers nerve fibers deteriorates and causes nerve damage. MS is most common in individuals of European descent; however, it still affects other populations. Previous research studies have found that the average age of onset for MS in ethnic minorities is 28.6 years, where those of European descent had an average of 32.8 years. Furthermore, studies have shown that non-Hispanic black patients with MS died at an earlier age and have an increasing mortality trend in comparison to white patients with MS. This suggests that MS takes a different toll on individuals depending on race; understanding this difference can aid us in better understanding MS onset and how to better accommodate ethnic minorities with MS. We acquired de-identified electronic health records (EHRs) from Vanderbilt University Medical Center BioVU that were genotyped on the Illumina MEGAex platform. We analyzed age of onset of MS in black and white individuals diagnosed with MS in an aim to identify single-nucleotide polymorphisms (SNPs) associated with earlier onset of MS in Black individuals. After completing quality control (QC), we identified 184 non-Hispanic black individuals and 1540 non-Hispanic white individuals for analysis. Linear regression will be used to identify interactions between known MS risk SNPs associated with age of onset in white and black populations.

PrgmNr 3590 - Bayesian analyses of variant-specific features enable penetrance estimates across cardiac channelopathies

View session detail

Author Block: **M. O'Neill**[1], K. Kozek[1], L. Sala[2], Y. Wada[3], I. Denjoy[4], C. Egly[3], N. Makita[5], T. ISHIKAWA[5], L. Crotti[2], C. Spazzolini[2], M-C. Kotta[2], F. Dagradi[2], S. Castelletti[2], M. Pedrazzini[2], M. Gnecchi[6], A. Leenhardt[4], J-E. Salem[7], S. Ohno[8], A. Glazer[3], J. D. Mosley[3], B. Knollmann[3], J. Blume[1], M. Horie[8], F. Extramiana[4], P. Schwartz[2], D. Roden[3], B. Kroncke[3]; [1]Vanderbilt Univ., Nashville, TN, [2]Istituto Auxologico Italiano IRCCS, Milan, Italy, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]AP-HP HÃ´pital Bichat, Paris, France, [5]Natl. Cerebral and Cardiovascular Ctr., Suita, Japan, [6]Univ. of Pavia, Pavia, Italy, [7]AP-HP PitiÃ©-SalpÃªtriÃ¨re Hosp., Paris, France, [8]Shiga Univ. of Med. Sci., Otsu, Japan

Disclosure Block: **M. O'Neill:** None.

The inability to accurately assess disease risk across carriers of rare variants in cardiac channelopathy genes represents a major impediment to personalized medicine. A common way to adjudicate these variants has been observational, commonly invoking case-control designs. While representing a powerful approach, such methods rely *a priori* on clinical observations which suffer from strong ascertainment bias and heterogenous presentations of variant carriers; the clinical implications for variants can vary strikingly across individuals even when evidence of variant pathogenicity is strong. Sequencing efforts have revealed both a high number of ultra-rare variants in healthy individuals and a much higher burden of rare, definitively annotated disease-causing variants than is compatible with the individual disease prevalence. We propose an inverse model for variant interpretation, whereby variant-specific features inform a disease penetrance prior in a Bayesian framework. Herein we present penetrance estimates for long QT syndrome 1-3 (LQT1-3) and Brugada Syndrome (BrS) based on literature curation of phenotyped individuals, *in silico* predictors, and protein structure-derived features for variants in the causative genes *KCNQ1* (LQT1), *KCNH2* (LQT2), and *SCN5A* (BrS + LQT3). Our results show that penetrance of pathogenic BrS variants is approximately 50%, while those of LQTS cluster near 80%, consistent with a threshold model by which total compensation against an insult from an LQTS-associated variant is less than that for BrS-associated variants. We provide model validation for two genotype-phenotype relationships on an international cohort of 1610 *KCNQ1* variant carriers and 933 *KCNH2* variant carriers. Furthermore, a remarkable heterogeneity of individual residue penetrance can be appreciated across traditionally annotated â□□pathogenic domainsâ□□. Through this framework we found that the innate diagnostic information one learns about a variant from three-dimensional variant location, *in vitro* data, and *in silico* predictors is equivalent to the diagnostic information one obtains by clinically phenotyping 10-20 heterozygous carriers. Altogether, we provide a powerful means by which to assess variant penetrance and inform clinical decision making prospectively, before the observation of an affected or unaffected heterozygous carrier.

PrgmNr 3591 - BrainXcan identifies brain features associated with Alzheimer's disease using large scale genetic and neuroimaging data

View session detail

Author Block: **E. Wu**[1], Y. Liang[1], O. Melia[1], T. S. Brettin[2], A. Brown[3], H. Im[1]; [1]The Univ. of Chicago, Chicago, IL, [2]Argonne Natl. Lab., Lemont, IL, [3]Univ. of Dundee, Dundee, United Kingdom
Disclosure Block: **E. Wu:** None.
Alzheimerâ□□s disease (AD) is a progressive neurodegenerative disease with an enormous health burden on individuals and society with no effective treatment to date. Genome-wide association studies (GWAS) have discovered 75 loci associated with risk of AD implicating amyloid and tau metabolism, endocytosis, innate immunity pathways. However, most loci reside in non-coding regions of the genome, making it difficult to understand their mechanism of action.
Brain magnetic resonance imaging (MRI) is a promising technique for diagnosis and examination of AD etiology and pathology. Previous studies have been successful in identifying whole-brain and temporal lobe, particularly hippocampal, atrophy as powerful indicators of Alzheimerâ□□s neurodegeneration. However, low statistical power and lack of standardized data analysis limit the reliability of these studies. Whether these markers are causal or emerge as a result of disease progression is also unclear without additional analyses.
Here we apply a newly developed method, BrainXcan, which brings together neuroimaging genomic toolsets to identify structural features of the brain involved in the pathogenesis and progression of AD. This allows for discovery of causal features implicated in AD by using prediction weights trained with the uniformly processed UK Biobank's brain MRI image-derived phenotypes and correlating the genetically predicted brain features with disease status.
Applying BrainXcan to Alzheimerâ□□s disease using T1-weighted and diffusion MRI data revealed 11 significant associations among 261 image-derived phenotypes, including increased gray matter volumes in the occipital fusiform gyrus and caudate, as well as features indicating reduced connectivity and white matter tissue integrity consistent with our understanding of neurodegeneration. However, the absence of significant temporal lobe atrophy and other characteristic structural features identified in previous studies warrants further investigation into the genetic basis of these features and the direction of causality in the context of Alzheimerâ□□s disease. Taken together, our results highlight the promise of BrainXcan as a powerful tool to dissect the biology of AD and add new lines of evidence for existing results.

PrgmNr 3592 - Circulating proteins to predict adverse COVID-19 outcomes

View session detail

Author Block: **C-Y. Su**[1,2], S. Zhou[1,2], E. Gonzalez-Kozlova[3], G. Butler-Laporte[1,2], E. Brunet-Ratnasingham[4], T. Nakanishi[1,2], W. Jeon[1], D. Morrison[2], L. Laurent[2], N. Kimchi[2], The Mount Sinai COVID-19 Biobank Team, J. Pineau[1], V. Mooser[1], N. D. Beckmann[3], E. Kenigsberg[3], E. Schadt[3], S. Kim-schulze[3], A. W. Charney[3], S. Gnjatic[3], D. E. Kaufmann[4], M. Merad[3], J. B. Richards[1,2]; [1]McGill Univ., Montreal, QC, Canada, [2]Lady Davis Inst., Jewish Gen. Hosp., Montreal, QC, Canada, [3]Icahn Sch. of Med. at Mount Sinai, New York, NY, [4]Ctr. Hosp.ier de lâ□□UniversitÃ© de MontrÃ©al, Montreal, QC, Canada
Disclosure Block: **C. Su:** None.
Predicting COVID-19 severity is difficult, and the biological pathways involved are not fully understood. To approach this problem, we measured 4,701 circulating human protein abundances in two independent cohorts totaling 986 individuals. We then trained prediction models including protein abundances and clinical risk factors to predict adverse COVID-19 outcomes in 417 subjects and tested these models in a separate cohort of 569 individuals. For severe COVID-19, a baseline model including age and sex provided an area under the receiver operator curve (AUC) of 65% in the test cohort. Selecting 92 proteins from the 4,701 unique protein abundances improved the AUC to 88% in the training cohort, which remained relatively stable in the testing cohort at 86%, suggesting stability of the model. Proteins selected from different adverse COVID-19 outcomes were enriched for cytokine and chemokine receptors, but half of the enriched pathways were not immune-related. Taken together, these findings suggest that circulating proteins measured at early stages of disease progression are reasonably accurate predictors of adverse COVID-19 outcomes. Further research is needed to understand how to incorporate protein measurement into clinical care.

PrgmNr 3593 - Detecting *DHCR7* variants consistent with Smith-Lemli-Opitz syndrome in autism patients

View session detail

Author Block: **J. B. Cordero**[1], R. Y. Cordero[1], E. Tierney[2], F. D. Porter[3], C. A. Wassif[4], C. L. Simpson[5]; [1]Genetics, Genomics, and Informatics, Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [2]Kennedy Krieger Inst, Baltimore, MD, [3]NIH, Bethesda, MD, [4]NICHD/HDB/NIH, Bethesda, MD, [5]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN

Disclosure Block: **J.B. Cordero:** None.

Introduction: *DHCR7* encodes the enzyme that catalyzes the conversion of 7-dehydrocholesterol (7-DHC) to cholesterol. Defective cholesterol synthesis results in low cholesterol and increased concentration of 7-DHC. Mutations in *DHCR7* cause Smith-Lemli-Opitz syndrome (SLOS). Most SLOS patients are compound heterozygotes and present with congenital abnormalities, intellectual disability, and autism. Over 140 different mutations in *DHCR7* have been reported. In this study, we identified the prevalence of individuals with known *DHCR7* variants in MSSNG, an autism genomic database.

Methods: We examined whole genome sequencing (WGS) *DHCR7* variant data from MSSNG subjects with autism and identified individuals with â¥ 2 SLOS *DHCR7* known pathogenic variants. We analyzed lipid and behavioral phenotypic data in subsets of subjects in relation to the WGS analysis results. The frequency of identified individuals with the SLOS *DHCR7* pathogenic variants was then compared with control unaffected family members.

Results: A total of 7187 MSSNG subjects (3425 affected; 3762 unaffected) from 2756 families were included in the study, majority (80%) of which belonged to family trios or quads. Preliminary results from a subset of MSSNG database identified Autism-Genetic-Resource-Exchange (AGRE) subjects as having potentially deleterious *DHCR7* mutations. Analysis for the remaining MSSNG subjects is ongoing.

Conclusion: Early recognition and management of autism patients with SLOS-related *DHCR7* variants can help improve the course of the disorder. Further, the study of autism patients with concomitant *DHCR7* mutations affecting cholesterol synthesis may help elucidate pathways and relationships of autism and lipids.

PrgmNr 3594 - Dissecting the causal relationships between obesity and multiple sclerosis risk

View session detail

Author Block: **E. Misicka**, F. B. Briggs; Case Western Reserve Univ. Sch. of Med., Cleveland, OH
Disclosure Block: **E. Misicka:** None.
Background. Multiple sclerosis (MS) is an autoimmune disease of the central nervous system. Observational studies have demonstrated that higher body-mass index across the lifespan confers MS risk, and genetic instrumental variable (GIV) analyses have confirmed these relationships. How BMI contributes to MS risk is not clear, and it is unknown if other obesity-related traits are also independent risk factors for MS.
Objectives. To examine the causal effects of multiple measures of obesity on MS risk in non-Hispanic whites using genetic instrumental variable analyses.
Methods. Two-sample Mendelian randomization (2S-MR) is a statistical approach that uses summary statistics from genetic association studies to examine causal relationships between multiple phenotypes. Genome-wide association (GWA) summary statistics were obtained for BMI (N=681,275), waist-hip ratio (WHR; N=694,649), visceral adipose tissue (VAT; N=325,123), and the ratio of arm, leg, and trunk fat to total body fat (AFR, LFR, and TFR; N=362,499). GWA summary statistics for MS risk were available for 14,802 cases and 26,703 controls. GIVs for each phenotype were variants with associations that met genome-wide significance (p-8) in the respective studies. GIVs were linkage disequilibrium pruned (r^2**Results.** BMI and VAT were significantly associated with MS risk ($\hat{I}^2_{BMI}$=0.25, p_{BMI}VAT=0.24, p_{VAT}=0.043), and associations persist when removing pleiotropic GIVs ($\hat{I}^2_{BMI}$=0.25, p_{BMI}VAT=0.35, p_{VAT}BMI=0.19, p_{BMI}=0.007). However, VAT did not appear to be associated with MS Risk when accounting for BMI ($\hat{I}^2_{VAT}$=0.09, p_{VAT}=0.24). Mediation analysis did not demonstrate strong reciprocal mediating relationships between VAT and BMI in their relationships with MS risk.
Conclusions. Measures of obesity relate differently to MS risk, suggesting that various forms of body fat may play different roles in MS pathogenesis, and merits further investigation.

PrgmNr 3595 - Evaluation of genetic risk scores for incisional surgeries in primary open-angle glaucoma patients

View session detail

Author Block: **A. R. Waksmunski**[1,2], C. L. Nealon[3], T. G. Kinzy[1,2,3], C. W. Halladay[4], P. Simpson[3], R. L. Canania[3], S. A. Anthony[3], D. P. Roncone[3], L. Sawicki-Rogers[5], J. Leber[5], J. Dougherty[5], P. B. Greenberg[6,7], D. C. Crawford[1,2,3], S. K. Iyengar[2,3], J. M. Sullivan[8], W-C. Wu[9], N. S. Peachey[3,10,11], J. N. Cooke Bailey[1,2,3]; [1]Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, [2]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [3]Cleveland VA Med. Ctr., Cleveland, OH, [4]Ctr. of Innovation in Long Term Services and Supports, Providence VAMC, Providence, RI, [5]VA Western NY Hlth.care System, Buffalo, NY, [6]Section of Ophthalmology, Providence VAMC, Providence, RI, [7]Div. of Ophthalmology, Alpert Med. Sch., Brown Univ., Providence, RI, [8]Res. Service, VA Western NY Hlth.care System, Buffalo, NY, [9]Section of Cardiology, Med. Service, Providence VA Med. Ctr., Providence, RI, [10]Cole Eye Inst., Cleveland Clinic Fndn., Cleveland, OH, [11]Dept. of Ophthalmology, Cleveland Clinic Lerner Coll. of Med. of Case Western Reserve Univ., Cleveland, OH

Disclosure Block: **A.R. Waksmunski:** None.

Primary open-angle glaucoma (POAG) is a complex, degenerative eye disease for which early treatment intervention can mitigate severe visual impairment and blindness. To date, over 125 loci have been associated with POAG, but most of the variants from these loci have small effects and do not individually predict the need for certain POAG treatments. Rather than considering contributions of individual variants, the cumulative impact of these variants can be evaluated via a genetic risk score (GRS). To assess the clinical utility of a POAG GRS for predicting the need for invasive treatments, we tested a GRS based on 127 POAG risk variants from the literature in POAG patients enrolled in the Veterans Affairs (VA) Million Veteran Program (MVP), a large electronic health record (EHR)-linked biobank, who have and have not undergone incisional glaucoma surgeries. Unweighted and weighted GRS were calculated for 5,830 MVP POAG study participants with imputed genetic data and harmonized ancestry and race/ethnicity data, including 2,448 Black or African American (AA; cases=223) and 3,382 white or European American (EA; cases=222) Veterans. Weighted GRS calculations utilized effect estimates reported from published (i) cross-ancestry meta-analyses and (ii) ancestry-specific analyses. For our association analyses, MVP POAG patients with EHR codes for incisional glaucoma surgeries were sub-categorized into 3 case groups based on procedures they had: (1) trabeculectomy, (2) glaucoma drainage implants (GDI), or (3) both procedures. MVP POAG patients without any EHR codes for these procedures were considered controls. We performed association analyses of the weighted and unweighted 127-variant GRS with binary surgery phenotypes via logistic regression using unadjusted models as well as models adjusting for age, sex, and 5 sample-specific principal components. Receiver operating characteristic (ROC) curves were created for all models and statistically compared to evaluate GRS performance. In EA, weighted and unweighted GRS were both significantly associated with GDI. In AA, GRS weighted with ancestry-specific and cross-ancestry meta-analysis effect estimates were significantly associated with trabeculectomies. Ancestry-specific GRS were significantly associated in EA and AA who had undergone both trabeculectomies and GDI. For all three case groups, GRS using ancestry-specific weights outperformed the other GRS approaches we evaluated, suggesting that ancestry-specific GRS better classified patients requiring these invasive treatments. Future studies will aim to develop GRS to predict POAG risk stratification for any type of POAG treatment.

PrgmNr 3596 - Genetic Variation in Genes Associated with GERD in COPD

View session detail

Author Block: **A. Wilson**[1], P. Lakshman Kumar[2], M. Cho[3], B. D. Hobbs[4], H. Tiwari[1], G. Washko[5], S. Lutz[6], C. P. Hersh[7], E. Silverman[3], M-L. McDonald[1]; [1]Univ. of Alabama at Birmingham, Birmingham, AL, [2]Birmingham, AL, [3]Channing Div. of Network Med., Boston, MA, [4]Needham, MA, [5]Div. of Pulmonary and Critical Care Med., Boston, MA, [6]Dept. of Population Med., Boston, MA, [7]Canton, MA
Disclosure Block: **A. Wilson:** None.
Chronic Obstructive Pulmonary Disease (COPD) is one of the leading causes of death globally.Gastroesophageal reflux disease (GERD) is a prevalent (as high as 78%) comorbidity in COPD, and is associated with worse pulmonary symptoms, reduced quality of life, and increased exacerbations. We tested the hypothesis genetic variation is associated with GERD in COPD using a genome-wide association study (GWAS) approach utilizing the BioData Catalyst platform. Stratified, single variant analyses were performed using whole genome sequence data from TopMed in participants with COPD and GERD among African American (AA) (N= 822) and Non-Hispanic White (NHW) (N= 2,842) participants from the COPDGene Study. Significance was defined as p-8. In the NHW population, 6 variants were significantly associated with GERD in COPD. Two variants on chromosome 4 (4:187223386 and 4:187241021) are intergenic to the *AC108046.1* and *AC097652.1* novel transcripts which have been identified as RNA genes of the lncRNA class. Four variants on chromosome 2 (2:29893698, 2:29894085, 2:29896530, and 2:29894808) are intragenic to the *ALK* gene. *ALK*, ALK Receptor Tyrosine Kinase, encodes a receptor tyrosine kinase, which belongs to the insulin receptor superfamily. Among AA participants with COPD, several variants were significantly associated with GERD in COPD. The top variant associated with GERD in COPD was 1:5620209, which is intragenic to *AL365255.1. AL365255.1* is a novel transcript which has been found to be an RNA gene. Another variant significantly associated with GERD in COPD was 1:71792134, which was intragenic to the *NEGR1* gene. *NEGR1* encodes the neuronal growth regulator 1 protein, which is involved in cell adhesion. In conclusion, a number of significant single variants were identified in both NHW and AA populations, which warrant further exploration for their potential utility as biomarkers or therapeutic targets of GERD in COPD.

PrgmNr 3597 - Genome-wide survival analysis of dental caries incidence

View session detail

Author Block: **T. Zou**[1], K. Neiswanger[2], E. Feingold[1,3], B. Foxman[4], D. McNeil[5,6], M. L. Marazita[1,2], J. R. Shaffer[1,2]; [1]Dept. of Human Genetics, Graduate Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [2]Ctr. for Craniofacial and Dental Genetics, Dept. of Oral and Craniofacial Sci., Sch. of Dental Med., Univ. of Pittsburgh, Pittsburgh, PA, [3]Dept. of Biostatistics, Graduate Sch. of Publ. Hlth., Univ. of Pittsburgh, Pittsburgh, PA, [4]Dept. of Epidemiology, Univ. of Michigan Med. Sch., Ann Arbor, MI, [5]Dept. of Psychology, Eberly Coll. of Arts and Sci., West Virginia Univ., Morgantown, WV, [6]Dept. of Dental Practice & Rural Hlth., Sch. of Dentistry, West Virginia Univ., Morgantown, WV
Disclosure Block: **T. Zou:** None.
Objectives: The incidence of dental caries can be influenced by multiple factors including behavioral, environmental and genetic components. The aim of this study was to use survival analysis to identify the potential risk factors and genetic variants associated with dental caries incidence in a birth cohort.**Methods:** The Center for Oral Health Research in Appalachia, cohort 2 (COHRA2) recruited and prospectively followed pregnant women and their children starting in 2011. A total of 909 children were included in this study; each was followed annually from birth with 7 years being the longest follow-up to date in this ongoing study. Annual intra-oral examinations were performed to assess dental caries experience, including the approximate time to first carious lesion. Cox hazards models were used to assess the associations of time to event with self- reported risk factors and 4.9 million genetic variants ascertained using a genome-wide SNP array.**Results:** A total of 196 children (21.56%) had their first primary tooth caries event during the follow-up period. The average survival time was 3.23 years. Household income, home water source, motherâ□□s educational attainment, motherâ□□s tooth brushing frequency, motherâ□□s dental caries experience, breastfeeding status, and breastfeeding duration were individually associated with dental caries hazard in univariate models, while only household income, home water source, and motherâ□□s educational attainment were significantly associated in the multivariate model of all potential risk factors simultaneously. The heritability (i.e., proportion of variance explained by genetics) of the time to event trait was 54.4% (PConclusions: Our findings indicate that household income, motherâ□□s educational attainment, and home water source may be independently-operating risk factors for dental caries incidence, and that the time to event of first carious lesion is heritable. We nominate several suggestive loci for further investigation.

PrgmNr 3598 - GWAS of self-report and imputed phenotypes capture distinct but complementary genetic architecture of stuttering

View session detail

Author Block: **D. M. Shaw**[1], H. Polikowsky[1], D. Pruett[1], H-H. Chen[2], L. E. Petty[3], K. Viljoen[4], J. Beilby[4], 23andMe INC, R. Jones[1], S. Kraft[5], J. E. Below[3]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Vanderbilt Univ Med Ctr., Nashville, TN, [4]Curtin Univ., Perth, Australia, [5]Wayne State Univ., Detriot, MI

Disclosure Block: **D.M. Shaw:** None.

Developmental stuttering is a common speech disorder characterized by disruptions in the forward movements of speech. Aggregation of large clinically ascertained cases of developmental stuttering for genetic analysis is difficult as stuttering is not often treated in a hospital setting but dispersed through schools a private therapy clinics. Additionally, stuttering has a high rate of recovery in early childhood. These factors have resulted in limited population based-discovery of genetic risk factors for the disorder.To overcome recruitment challenges we assessed the genetic effects of two related phenotypes. First, we utilized a classification and regression tree model built using stuttering comorbidities as predictive features (PheML) to impute the stuttering phenotype in BioVU based on each subjectâ□□s phenotypic profile. Second, we utilized responses from a self-report questionnaire administered through 23andMe INC. Each method amassed a case set large enough to sufficiently power genome-wide association analyses with genome-wide significant signals. We utilized the summary statistics from each of these analyses to develop two separate polygenic risk score (PRS) models to measure genetic liability. Using an independent set of clinically confirmed stuttering cases and matched controls, we demonstrate that each model effectively measures stuttering liability. Confirmed stuttering cases scored significantly higher than matched controls for both the 23andMe PRS model ($p=5.81x10^{-28}$, AUC=0.61) and the PheML model ($p=6.83x10^{-39}$, AUC=0.60). However, the 23andMe model failed to show increased stuttering liability in our PheML stuttering sample set (p=0.08, AUC=0.49), indicating that these methods are capturing two distinct aspects of the overall phenotype.We developed a third PRS model that incorporated the summary statistics from both the 23andMe and PheML analyses, and outperformed each individual model. With this model, clinically validated stuttering cases showed significantly higher genetic liability for stuttering compared to controls ($p = 8.88x10^{-43}$), and the overall predictive ability of the model improved (AUC = 0.657 and AUC = 0.807 for subjects within the top and bottom 5% of liability scores).Using two models for developmental stuttering that capture distinct genetic signatures, we are able to build an integrated model that better predicts true developmental stuttering in the general population. The presence of multiple distinct genetic signatures within stuttering speaks to the polygenicity of this phenotype. By leveraging both signatures we are able to more completely capture true stuttering cases.

PrgmNr 3599 - Identification of Genetic Loci Associated with Cognitive Preservation in the Midwestern Amish

View session detail

Author Block: **L. R. Main**[1], Y. Song[2], R. A. Laux[1], K. Miskimen[3], M. D. Osterman[4], D. Dorfsman[5], M. L. Cuccaro[6], P. Ogrocki[1], A. J. Lerner[1], J. M. Vance[5], D. M. Fuzzell[1], S. Fuzzell[1], L. Caywood[7], M. Prough[8], L. D. Adams[9], J. Clouse[10], A. H. Lynn[1], S. Herington[11], W. K. Scott[5], M. A. Pericak-Vance[12], J. L. Haines[13]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Case Western Reserve Univ., Sch. of Med., Cleveland, OH, [3]Lakewood, OH, [4]Cleveland, OH, [5]Univ. of Miami, Miami, FL, [6]Univ Miami Sch Med, Miami, FL, [7]Coldwater, MI, [8]Univ. of Miami, Bristol, TN, [9]Univ Miami, Miami, FL, [10]Univ. of Miami, Niles, MI, [11]Univ. of Miami, MIller Sch. of Med., Hussman Inst. for Genomics, Elkhart, IN, [12]Miami, FL, [13]Case Western Reserve Univ, Cleveland, OH

Disclosure Block: **L.R. Main:** None.

Understanding the biological underpinnings of cognitive preservation despite advanced age is of growing interest as the world population ages. However, most studies of older individuals have focused on diseases, such as Alzheimer disease (AD), instead of on the preservation of cognitive function. To fill this gap, we have initiated studies of older individuals in the Midwestern Amish from Ohio and Indiana, a population that is isolated both genetically and culturally, practicing a conservative lifestyle and endogamy. Studies of the Amish allow us to better control for many confounding variables, and therefore examine more directly for genetic effects. We obtained genomic data on over 900 individuals who are 76 years or older and are judged cognitively unimpaired based on 3MS and CERAD neuropsychological exams. To help inform our analyses, we included genotypes of siblings and parents who are cognitively impaired. With extensive genealogical records, we connected all individuals through a pedigree with over 8,000 individuals across 15 generations. To extract maximal information from this complex pedigree, both a genome-wide association study (GWAS) and linkage analyses are being performed. Two-point nonparametric linkage is underway across all chromosomes using sub-pedigrees generated from the all-connecting pedigree. Overlapping sets of sub-pedigrees have been created by breaking the larger pedigree at different branch points, and by setting the size of sub-pedigrees to different thresholds. Linkage analysis will be performed multiple times across multiple overlapping sets of sub-pedigrees to assess robustness of findings. Average LOD scores will be taken from areas of interest across the different sub-pedigree formations. Regions of suggestive or significant loci will be followed up with multipoint and parametric linkage analyses in these locations. Our preliminary GWAS was completed with the GENESIS R package and identified six suggestive loci (P-5) on chromosomes 5 (rs13177199), 7 (rs140733969), 11 (exm2250129), 13 (rs681938), 15 (rs1347455), and 16 (rs12596728). None of these SNPs fall within previously identified regions associated with risk for AD, but the suggestive SNPs on chromosomes 11 and 16 are near known AD risk genes PICALM and IQCK, respectively. The genomic inflation factor of the GWAS is 0.98, indicating that there is slight over adjustment to the data, most likely due to the high degree of relatedness in the samples.

PrgmNr 3600 - Investigating White Blood Cell Count and Platelet Biology as a Possible Driver of the Comorbidity Between Coronary Artery Disease and Major Depressive Disorder

View session detail

Author Block: **K. Singh**[1], D. E. Gustavson[2], L. K. Davis[3]; [1]Vanderbilt Univ., Nashville, TN, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Vanderbilt Univ Med Ctr., Nashville, TN
Disclosure Block: **K. Singh:** None.
Background: Coronary Artery Disease (CAD) has a lifetime prevalence of 35% and is often comorbid with Major Depressive Disorder (MDD). Approximately 40% of people with either CAD or MDD will develop the other condition. Our previous work demonstrated that the polygenic risk score (PRS) for MDD was significantly associated with manifestations of CAD (OR, 1.10, p=1.08x10-5) in a large hospital population. We also identified elevated levels of inflammatory biomarkers including white blood cell and platelet counts in people with high genetic risk for MDD and CAD. Previous work also supports the role of inflammation in both CAD and MDD independently. Taken together, these studies indicate that genetic factors may predispose people to comorbid MDD with CAD, and that inflammation could be an important biological link between the conditions. Here, we have expanded this work to include both phenotypes and quantitative physiological measurements in a multivariate framework.Methods: Our research was conducted using BioVU. BioVU is Vanderbiltâ□□s collection of ~100,000 de-identified DNA samples from Vanderbilt hospitals. We employed LabWAS, a biomarker-based association method to identify labs associated with CAD and MDD comorbidity. We next performed multivariate and mediation analyses by applying Genomic Structural Equation Modeling (Genomic SEM) to the GWAS summary statistics for CAD, MDD, WBC count, and platelet count.Results: Using a Genomic SEM model, we found a single factor in which WBC count and platelet count explained 1% and 16% of the SNP-based genetic correlation between MDD and CAD, respectively. The mediation paths models using the summary statistics data show that WBC and platelets mediate approximately 1% and 13% of MDD-CAD comorbidity. Therefore, WBC and Platelets combined explain 14.4% of the total genetic correlation between MDD and CAD. Hence, these results suggest that genetically regulated WBC count and Platelet count may mediate the MDD-CAD comorbidity. Conclusions: Our findings provide further support for genetically regulated inflammation linking MDD and CAD, particularly WBC count and Platelet count. These results demonstrate that variance in comorbid MDD and CAD is, in part, controlled by genetic regulation of white blood cell and platelet count. We will also present results from current work including multivariate GWAS and expansion of these models to incorporate other inflammatory markers such as CRP, IL10RB and IL6.

PrgmNr 3601 - Joint intron splicing-based transcriptome-wide association study identifies new candidate susceptibility genes for breast cancer

View session detail

Author Block: **G. Gao**[1], P. Fiorica[1], A. N. Barbeira[2], J. McClellan[1], D. Huo[1], H. Im[2]; [1]Dept. of Publ. Hlth.Sci., Univ. of Chicago, Chicago, IL, [2]The Univ. of Chicago, Chicago, IL
Disclosure Block: **G. Gao:** None.

In this study, we proposed a joint intron splicing-based transcriptome-wide association study approach (IntronXcan) that combined information from multiple excised introns in a gene across multiple tissues. IntronXcan used splicing prediction models trained in 47 tissues in the GTEx (v8) data with a multivariate adaptive shrinkage (mash) method, which can jointly estimate effects of splicing quantitative trait loci (sQTLs) in multiple tissues, accounting for correlation among nonzero sQTL effects in different tissues. We applied IntronXcan to the GWAS summary statistics from the Breast Cancer Association Consortium (BCAC) of about 229,000 women of European ancestry. We identified 550 genes significantly (P -^{6}) associated with breast cancer. We replicated 75 of these 550 genes at significance level (P -^{5}) when applied IntronXcan to GWAS summary statistics from an independent dataset on breast cancer extracted from UK Biobank. To determine if these genes are independent of previously identified breast cancer loci, we performed a conditional and joint analysis (COJO) by conditioning summary statistics on previously published (risk) index SNPs within +/- 10 Mb of each significant gene. We then re-performed IntronXcan using the conditional summary statistics. Results showed that 31 of the 75 replicated genes were still significant. We treat these 31 genes as conditionally independent of known breast cancer GWAS loci. Of these 31 genes, 13 novel candidate susceptibility genes are located beyond 500kb from their nearest GWAS index SNP. We similarly performed gene expression-based joint TWAS (S-MultiXcan) that integrated predicted expression data from multiple tissues. We identified 311 genes significantly associated with breast cancer using BCAC summary statistics (P â□□6). Among them, 23 genes were replicated (P â□□5) in our S-MultiXcan analysis of the independent UKB dataset. Six of the 23 replicated genes remained significant after conditioning on previously identified index SNPs. We further performed a meta-analysis to combine BCAC and UKB GWAS summary statistics and then applied splicing-based IntronXcan and expression based S-MultiXcan to the meta-analysis summary statistics. We identified an additional 162 and 68 significant genes, respectively. Our analyses indicated that the splicing based IntronXcan could identify a greater number of significant genes than the gene expression-based S-MultiXcan and individual intron splicing-based TWAS.

PrgmNr 3602 - Overwhelming numbers of patients who carry pathogenic familial hypercholesterolemia variants are predicted to have low clinical risk in the UK Biobank

View session detail

Author Block: **J. Fife**[1], C. Cassa[2]; [1]Brigham and Women's Hosp., Cambridge, MA, [2]Harvard Med. Sch., Cambridge, MA
Disclosure Block: **J. Fife:** Salary/Employment; nference Inc.
Clinical risk assessment remains challenging even in established disease genes, as many variants are rare. We develop a model which integrates clinical risk factors with variant-level features to estimate clinical risk in nine disease genes, using 200,000 UK Biobank exomes. These risk estimates are highly concordant with clinical outcomes in patients who carry rare variants of uncertain significance (VUS). In genes associated with breast cancer (BC) and familial hypercholesterolemia (FH), we distinguish between individuals with elevated risk versus population-level disease risk (RR>3.81, logrank p

PrgmNr 3603 - Polygenic risk score analysis provides evidence of unique genetic architecture of Alzheimerâ□□s disease in an Amish population

View session detail

Author Block: **M. D. Osterman**[1], Y. Song[2], L. D. Adams[3], R. A. Laux[1], L. Caywood[4], M. Prough[5], J. Clouse[6], S. Herington[7], S. H. Slifer[8], A. H. Lynn[1], M. Fuzzell[1], S. L. Fuzzell[1], S. D. Miller[1], K. Miskimen[9], L. R. Main[10], D. Dorfsman[11], P. Ogrocki[1], A. J. Lerner[1], J. Ramos[12], J. M. Vance[11], M. L. Cuccaro[13], W. K. Scott[11], M. A. Pericak-Vance[14], J. L. Haines[15]; [1]Case Western Reserve Univ., Cleveland, OH, [2]Case Western Reserve Univ., Sch. of Med., Cleveland, OH, [3]Univ Miami, Miami, FL, [4]Coldwater, MI, [5]Univ. of Miami, Bristol, TN, [6]Univ. of Miami, Niles, MI, [7]Univ. of Miami, MIller Sch. of Med., Hussman Inst. for Genomics, Elkhart, IN, [8]Univ. of Miami Miller Sch. of Med., Miami, FL, [9]Lakewood, OH, [10]Case Western Reserve Univ., Lakewood, OH, [11]Univ. of Miami, Miami, FL, [12]Univ. of Miami, Miller Sch. of Med., Miami, FL, [13]Univ Miami Sch Med, Miami, FL, [14]Miami, FL, [15]Case Western Reserve Univ, Cleveland, OH
Disclosure Block: **M.D. Osterman:** None.
Alzheimerâ□□s disease (AD) is the most prevalent type of dementia and is currently estimated to affect 5.8 million Americans. Risk for AD is multifactorial, including both genetic and environmental risk factors. Much of AD genomic research has focused on identifying risk variants, primarily in European ancestry populations. After immigration to the United States, the Amish experienced a genetic bottleneck, making it likely that their underlying genetic architecture is different from the broader European ancestry population. Here, we compare the genetic risk for Alzheimerâ□□s disease in an Amish population to a general European ancestry population using polygenic risk scores (PRSs). Our study population is comprised of 1,981 adults recruited from Amish communities in Indiana and Ohio. Subjects were screened for cognitive impairment and further evaluated for AD and dementia. Genotype data were collected using Illumina genotyping chips. Imputation was performed based on a Haplotype Reference Consortium Panel. A non-Amish population of 2,470 adults recruited from clinics in Tennessee, North Carolina, and Florida was used as a comparison group. PRSs were generated using effect estimates of variants from the Jansen et al. (2019) genome-wide meta-analysis of AD weighted by their log-odds ratios and excluding variants within 500 kilobases of APOE. We compared the PRSs by source population and case status, after exclusion of individuals under age 75. Further, we evaluated their predictive ability on AD status with and without sex, age, and APOE genotype covariates. As expected, the results indicated that there exists less variation in PRSs among the Amish population. After standardization of PRSs to a Normal(0,1) distribution, the mean PRS of AD affected Amish (0.040) was lower ($p = 5 Ã□ 10^{-4}$) than the mean PRS of the non-Amish cases (0.324). The mean PRS of the unaffected Amish (0.066) was higher (p -16) than the non-Amish controls (-0.458). The PRS was able to differentiate between non-Amish cases and controls (p -16) but not Amish affection status (p = 0.7). We found that addition of a PRS to an AD prediction model including sex, age, and APOE genotype did not improve area under the curve (AUC) in the Amish but improved AUC in the non-Amish population from 0.765 to 0.812. These results suggest that the Amish have a different genetic architecture than a general European-ancestry population, manifesting itself in less overall variation but also a lower relative importance of APOE. Because the prevalence of dementia in the Amish is similar to the general European population, this suggests that the Amish may harbor unique AD genetic risk variants.

PrgmNr 3604 - Prioritizing causal risk factors for severe COVID-19: an exhaustive Mendelian randomization study

View session detail

Author Block: **Y. Sun**[1], J. Zhou[1,2], K. Ye[3,4]; [1]Dept. of Genetics, Univ. of Georgia, Athens, GA, [2]Sch. of Publ. Hlth., Shanghai Jiao Tong Univ. Sch. of Med., Shanghai, China, [3]Univ. of Georgia, Athens, GA, [4]Inst. of Bioinformatics, Univ. of Georgia, Athens, GA
Disclosure Block: **Y. Sun:** None.
Identifying causal risk factors for severe coronavirus disease 2019 (COVID-19) is critical for its prevention and treatment. Many associated pre-existing conditions and biomarkers have been reported, but these observational associations suffer from confounding and reverse causation. Here, we perform a large-scale two-sample Mendelian randomization (MR) analysis to evaluate the causal roles of many traits in severe COVID-19. Our results highlight multiple body mass index (BMI)-related traits as risk-increasing: BMI (OR: 1.89, 95% CI: 1.51-2.37), hip circumference (OR: 1.46, 1.15-1.85), and waist circumference (OR: 1.82, 1.36-2.43). Our multivariable MR analysis further suggests that the BMI-related effect might be driven by fat mass (OR: 1.63, 1.03-2.58), but not fat-free mass (OR: 1.00, 0.61-1.66). Several white blood cell counts are negatively associated with severe COVID-19, including those of neutrophils (OR: 0.76, 0.61-0.94), granulocytes (OR: 0.75, 0.601-0.93), and myeloid white blood cells (OR: 0.77, 0.62-0.96). Furthermore, some circulating proteins are associated with an increased risk of (e.g., zinc-alpha-2-glycoprotein) or protection from severe COVID-19 (e.g., interleukin-3 receptor subunit alpha). Our study suggests that fat mass and white blood cells may underlie the etiology of severe COVID-19. It also prioritizes potential risk and protective factors that could serve as drug targets and guide the effective protection of high-risk individuals.

PrgmNr 3605 - Rare Variants in Non-coding Regulatory Regions and Non-syndromic Orofacial clefts

View session detail

Author Block: **A. Alade**[1], C. Comnick[1], W. Awotoye[1], E. Zeng[1], C. Pendleton[1], A. Visel[2], K. Paraiso[3], A. Oladayo[4], P. Mossey[5], L. Gowans[6], M. A. Eshete[7], W. Adeyemo[8], T. Naicker[9], C. Adeleke[1], A. Petrin[4], J. Olotu[10], M. Hassan[1], J. Pape[1], P. Donkor[11], T. Busch[1], M. L. Marazita[12], A. A. Adeyemo[13], J. C. Murray[14], A. Butali[15]; [1]Univ. of Iowa, Iowa, IA, [2]Lawrence Berkeley Natl Lab., Berkeley, CA, [3]Lawrence Berkeley Natl. Lab, Berkeley, CA, [4]Univ. of Iowa, Iowa City, IA, [5], Univ. of Dundee, Dundee, United Kingdom, [6]Kwame Nkrumah Univ. of Sci. and Technology, Kumasi, Ghana, [7]Addis Ababa Univ., Addis Ababa, Ethiopia, [8]Univ. of Lagos, Lagos, Nigeria, [9]Univ. of Kwazulu natal, Durban, South Africa, [10]Univ. of Port Harcourt, Port Harcourt, Nigeria, [11]Kwame Nkiruma Univ., Kumasi, Ghana, [12]Univ Pittsburgh, Pittsburgh, PA, [13]NIH, Bethesda, MD, [14]U of Iowa, Iowa City, IA, [15]Univ Iowa, Iowa City, IA

Disclosure Block: **A. Alade:** None.

BACKGROUND Orofacial clefts are the most common congenital birth defects accounting for 65% of all head and neck deformities. To date, Genome Wide Association Studies (GWAS) have identified over 60 common risk loci for non-syndromic orofacial clefts (NSOFCs). Collectively, these loci only explain 20-30% of its heritability. Interestingly, a substantial number of these loci are in the non-coding regulatory regions of the genome. To explain the â□□missing heritabilityâ□□, we investigated putative craniofacial and ectodermal enhancer elements for rare variants with probable etiological implication in the pathogenesis of NSOFCs. **METHODS** We investigated a total of 3178 individuals; 1019 cases (394 had Cleft lip, 415 Cleft lip Â± palate, 205 Cleft palate only, and 5 unknown) and 2159 unrelated controls. Candidate craniofacial and ectodermal enhancer elements were identified from previous literatures and mouse models. A total of 58,449 regulatory regions were considered. Following filtration to include only regions with more than one rare (MAF RESULTS We found a suggestive significant association (p = 0.009, Bonferroni-corrected cut-off for statistical significance was p PRDM16 gene (Chr1:3148388-3149216) with all clefts combined. For the putative ectodermal enhancers, we found a near significant association (p = 3.83e-05 and p = 4.51e-05 respectively; Bonferroni-corrected cut-off for statistical significance was p CRK and *RREB1* genes (Chr17:1358515-1360443 and Chr6:7137252-7138277) with the cleft lip with or without palate phenotype. PRDM16 is a known cleft candidate gene and p300 ChIP-Seq data showed enhancer activity for this region, part of which overlaps with data from ATAC-Seq data in relevant craniofacial tissues. The RREB1 gene codes for a zinc finger transcription factor and deletion of the short arm of chromosome 6 involving this gene present with Craniofacial dysmorphism. *CRK*-knockout mice showed cleft palate phenotype. **CONCLUSION** Our findings suggest a role for rare variants within craniofacial enhancer regions in the complex etiology of NSOFCs.

PrgmNr 3606 - Sex-specific analyses in inflammatory bowel disease in African Americans reveal new loci on chromosomes 4, 5, 6, and 13

View session detail

Author Block: **C. L. Simpson**[1], D. P. McGovern[2], S. Kugathasan[3], T. Haritunians[4], S. R. Brant[5]; [1]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [2]Los Angeles, CA, [3]Emory Univ Sch. of Med., Atlanta, GA, [4]Cedars Sinai Med. Ctr., Los Angeles, CA, [5]Rutgers Robert Wood Johnson Med. Sch., New Brunswick, NJ
Disclosure Block: **C.L. Simpson:** None.
Inflammatory bowel disease (IBD) is a leading cause of morbidity in the US and increasing in prevalence in all populations, including African Americans (AA). Most genetic loci have been identified in European populations, whereas other ethnicities have been underrepresented. IBD consists of 2 main subtypes, Crohnâ□□s Disease (CD) and Ulcerative Colitis (UC). CD is characterized by transmural inflammation and can affect any part of the gastrointestinal tract. UC is typically restricted to the mucosa and only affects the colon. The incidence and prevalence of IBD are different in men and women, with prepubertal girls at lower risk of CD, but post-puberty risk is higher in females. UC rates are similar in males and females under 45, with increased risk for men over 45. Here we present sex-specific and sex-interaction analyses of an AA IBD cohort. Genomewide SNP microarray data from 2345 AA cases of IBD (1646 CD, 583 UC, 116 unspecified IBD) and 5002 controls (derived from Health and Retirement Study and Kaiser Permanente cohorts) were used in sex-stratified, sex-adjusted, and sex-interaction multivariate logistic regression models. P values â□¤ $5x10^{-8}$ were considered genomewide significant (GWS). Integrating sex into the model showed novel GWS signals on 5q23.2 (rs73782531, P=$5.16x10^{-11}$, OR=0.077) and 6p23 (rs731187, P=$2.53x10^{-8}$,OR=0.67) for IBD, both in gene deserts. We found female-specific novel associations with UC on 4q21.21 (rs140502305, P=3.65 x 10^{-8}, OR=3.86), near PRKG2, which regulates intestinal secretion, and a signal on 13q32.1 (rs9584460, P=2.59 x 10^{-9}, OR=2.05) intronic to OXGR1 that interacts with leukotrienes associated with inflammation. A novel male-specific peak was GWS at 12q24.13 (rs3782429, P=2.24 x 10^{-8}, OR=5.69), intronic to RBM19, related to digestive system development in animal models. A significant peak in the HLA region of chromosome 6, long associated with UC, was more significant in males (rs9271737, P=$2.31x10^{-10}$, OR=0.41) than females was (rs9276518 P=$3x10^{-8}$, OR=1.77). In addition, a GWS signal on chromosome 9 intronic to TNC (rs115542730, P=$4.15x10^{-8}$, OR=1.92) was previously identified in our AA GWAS study and expression of which has been shown to be increased in IBD. No interaction effects were observed at the GWS level. Analysis of the X chromosome was also not GWS. Here we identify novel candidate genes in an AA population. Including sex in GWAS models and looking for sex-specific associations in traits with known sex bias can reveal new loci and provide evidence for asymmetric effect sizes in known loci. Further research is required to replicate our findings.

PrgmNr 3607 - The Smoking Independent Causal Effects of Alcohol and Socioeconomic Traits on Lung Cancer Development

View session detail

Author Block: **R. Pettit**[1], J. Byun[1], Y. Han[1], Q. T. Ostrom[2], K. M. Walsh[2], M. L. Bondy[3], R. Hung[4], J. D. McKay[5], C. I. Amos[1]; [1]Baylor Coll. of Med., Houston, TX, [2]Duke Univ., Durham, NC, [3]Stanford, Stanford, CA, [4]Lunenfeld-Tanenbaum Res. Inst., Sinai Hlth.System, Toronto, ON, Canada, [5]Intl. Agency for Res. on Cancer - World Hlth.Organization, Lyon, France
Disclosure Block: **R. Pettit:** None.
Introduction: Lung cancer is heritable and multiple traits have been associated with increased lung cancer susceptibility. Mendelian Randomization (MR) is an epidemiological method to interrogate such associations, estimating causal trait effects using single nucleotide polymorphisms as genetic instruments. MR methods can estimate causal effects independent of secondary exposures, such as smoking, and avoids confounding or reverse causality biases. In our work, we set out to characterize smoking-dependent and independent causal effects of trait lung cancer relationships including â□□educational ascertainmentâ□□ and â□□alcohol useâ□□. **Methods:** To estimate the causal effects of alcohol use and educational ascertainment on lung cancer (LC) we employed univariate and multivariate two-stage mendelian randomization (MVMR) techniques utilizing genome-wide association study summary statistic data from the United Kingdom Biobank (UKBB) and the TRICL-OncoArray LC consortium. We tested several UKBB education and alcohol-related traits for causal effects on overall LC and its histological subtypes. We tested for trait - LC effect modulation via confounding or mediation from smoking-related traits, including smoking â□□age of initiationâ□□, â□□cigarettes per dayâ□□, and â□□smoking cessationâ□□. Using MR and MVMR effect estimation methods, including inverse variance weighting (IVW), weighted median, weighted mode, and mr-egger, we found that alcohol and educational ascertainment traits had direct causal effects on lung cancer development independent of smoking related traits. **Results:** Using the IVW MR method, having no higher education or â□□Qualifications: None of the aboveâ□□ in the UKBB had 5.94 times increased odds of overall lung cancer (95% CI 3.39, 10.39, $p = 4.49x10^{-10}$, 62 SNPs), while â□□Qualifications: College or University degreeâ□□ had a 0.31 decreased likelihood (95% CI 0.23, 0.42, $p = 1.40x10^{-13}$, 169 SNPs). Further, the trait â□□Average weekly beer plus cider intakeâ□□ had an OR of 3.48 (95% CI 2.15, 5.64, $p = 4.08x10^{-7}$, 19 SNPs) with overall lung cancer risk. â□□Alcohol usually taken with mealsâ□□ however had a OR 0.19 (95% CI 0.094, 0.36, $p = 1.06x10^{-6}$, 30 SNPs) with overall lung cancer risk. These trends were modified but remained despite MVMR contingent modeling with smoking-related traits. **Conclusion:** â□□Educational ascertainmentâ□□ and â□□alcohol useâ□□ traits had significant causal relationships with lung cancer development independent of smoking. Characterizing these trait relationships allows for actionable clinician-patient decision-making and positions these traits as candidates to be incorporated in future lung cancer polygenic risk scores.

PrgmNr 3608 - Todayâ□□s genetic risk scores for primary open-angle glaucoma underperform in African-descent participants from the Veteran Administrationâ□□s Million Veteran Program (MVP)

View session detail

Author Block: **L. Cruz**[1,2], C. L. Nealon[3], T. G. Kinzy[3,1,2], C. W. Halladay[4], P. Simpson[3], R. L. Canania[3], S. A. Anthony[3], D. P. Roncone[3], L. R. Sawicki Rogers[5], J. N. Leber[5], J. M. Dougherty[5], P. B. Greenberg[6,7], J. M. Sullivan[5], W-C. Wu[8], S. K. Iyengar[3,2], D. C. Crawford[3,1,2], N. S. Peachey[3,9,10], J. N. Cooke Bailey[3,1,2]; [1]Cleveland Inst. for Computational Biology, Case Western Reserve Univ., Cleveland, OH, [2]Dept. of Population and Quantitative Hlth.Sci., Case Western Reserve Univ., Cleveland, OH, [3]VA Northeast Ohio Hlth.care Systems, Cleveland, OH, [4]Ctr. for Innovation in Long Term Services and Supports, Providence VAMC, Providence, RI, [5]Ophthalmology Section, VA Western NY Hlth.Care System, Buffalo, NY, [6]Ophthalmology Section, Providence VA Med. Ctr., Providence, RI, [7]Div. of Ophthalmology, Alpert Med. Sch., Brown Univ., Providence, RI, [8]Cardiology Section, Med. Service, Providence VA Med. Ctr., Providence, RI, [9]Cole Eye Inst., Cleveland Clinic Fndn., Cleveland, OH, [10]Dept. of Ophthalmology, Cleveland Clinic Lerner Coll. of Med., Case Western Reserve Univ., Cleveland, OH
Disclosure Block: **L. Cruz:** None.

Genetic risk scores (GRS) for primary open-angle glaucoma (POAG) have strong clinical utility potential. While POAG GRS associate with case status, diagnosis age, and disease-specific outcomes in European-descent populations, data are limited for African-descent populations despite their higher POAG burden (earlier onset, higher prevalence, increased likelihood of vision loss/blindness). To assess the trans-population performance of POAG GRS based on currently published risk variants, we accessed the diverse Million Veteran Program (MVP), a U.S. Department of Veterans Affairs-sponsored observational cohort study and mega-biobank with electronic health (EHRs) and health survey data currently available for >825,000 Veterans; >450,000 have genome-wide data available. We applied a validated computable phenotyping algorithm to MVP participantsâ□□ EHRs and identified 3,382 and 2,448 POAG cases and 58,832 and 5,665 controls representing European- and African-descent participants, respectively. For each MVP participant, GRS were calculated by summing 127 GWAS-associated POAG risk alleles from the literature as of 2021. Association analyses of unweighted and effect estimate-weighted GRS and binary POAG phenotype were performed via logistic regression unadjusted and adjusted for age, sex, and 10 sample-specific principal components. GRS score deciles and quartiles were tested for association via logistic regression, and receiver operating characteristic (ROC) curves were composed and statistically compared. GRS were significantly associated with POAG in unadjusted and adjusted models at p

PrgmNr 3609 - Transcriptome-wide association studies identify genetic determinants of brain-imaging phenotypes

View session detail

Author Block: **X. Bledsoe**[1], A. Hale[2], D. Zhou[1], E. Gamazon[3]; [1]Vanderbilt Univ., Nashville, TN, [2]Nashville, TN, [3]VUMC Clare Hall, Univ. of Cambridge, Nashville, TN
Disclosure Block: **X. Bledsoe:** None.
Advances in neuroimaging have facilitated an increasingly sophisticated understanding of brain circuitry, morphology, and connectivity. However, genetic determinants of brain structure and function are largely unknown. To understand the genetic basis of these traits, we perform transcriptome-wide association studies (TWAS) using Joint Tissue Imputation (JTI)- PrediXcan for 3,144 neuroimaging-derived phenotypes (NIDP). The resulting associations represent genetically-regulated gene expression (GREx) and GWAS phenotypes. Due to variation in GREx across different tissues, we conducted our analyses using 17 tissue-expression models from GTEx which include 13 major brain regions as well as the adrenal gland, sigmoid colon, transverse colon, and whole blood. Using a strict Bonferroni study-wide significance threshold, we identified 2,359,962 tissue-GREx pairs across 3,144 NIDPs. For select traits, we provide systematic replication using two independent datasets (ENIGMA). The plurality of GREx-NIDP associations are specific to a single tissue model out of the 17 different models included in the analysis. Gene ontology (GO) analysis demonstrates an enrichment in DNA maintenance pathways for GREx genes that are associated with NIDPâ□□s across all measured tissues whereas enrichment in synapse-related pathways for the GREx-NIDP associations are only observed in a single tissue. GO pathways involved in GREx-NIDP associations unique to brain tissue favored cell-cell interactions while those involved in non-brain tissues favored signaling moieties such as hormones, steroids, and chemokine elements. We demonstrate the utility of the GREx-NIDP resource through application to large-scale GWAS meta-analyses of schizophrenia and epilepsy.

PrgmNr 3611 - Utilizing PrediXcan to detect Crohn's disease susceptibility genes in African Americans

View session detail

Author Block: **R. Y. Cordero**[1], S. Kugathasan[2], D. P. McGovern[3], S. R. Brant[4], C. L. Simpson[1]; [1]Univ. of Tennessee Hlth.Sci. Ctr., Memphis, TN, [2]Emory Univ Sch. of Med., Atlanta, GA, [3]Los Angeles, CA, [4]Rutgers Robert Wood Johnson Med. Sch., New Brunswick, NJ
Disclosure Block: **R.Y. Cordero:** None.
Multiple genetic risk loci associated with inflammatory bowel disease (IBD) have been identified through genome-wide association studies (GWAS). However, very large sample sizes are essential in GWAS to increase statistical power to detect disease-related loci. The power to detect IBD loci in African Americans (AA) and ultimately identify risk genes has been limited by modest sample sizes. To overcome this challenge, gene-based association methods such as PrediXcan were developed to detect the relationship between genes and traits to reduce the multiple testing burden. We utilized PrediXcan to integrate expression quantitative trait loci (eQTL) from the Genotype-Tissue Expression (GTEx) study and IBD GWAS summary statistics using independent case-control datasets, totaling 843 Crohnâ□□s disease (CD) cases and 1678 controls from unrelated, self-identified AA individuals. An elastic net model in whole blood, transverse colon, small intestine, adipose tissue were used as part of our transcriptome imputation. Our initial results reveal 48 significant associations with CD, with the strongest association coming from *SPDYE6* (p, $4.63x10^{-8}$). Other top associations include *OIT3, BIVM, AC245041.1, HROB, CSNK1G1, ZNF488, RASA4, FAHD2B, LOC102724488,* and *TMEM106A.* We found significant gene associations (*C15orf61, FBXW8, FAM219B, CYP1A1, PARG, ASAH2, SCAMP2, MPI, RPP25, WARS2, KNOP1,* and *PPCDC*) in regions in our AA CD published GWAS (Brant, Simpson, Okou, et al, 2017). Four other loci (*DHX58, MAP3K8, IFI35,* and *RPS6KL1*) were within a megabase of established European loci. Here, we demonstrate how gene-based methods, informed by other â□□omics data, can improve our ability to detect known and novel genes associated with CD in AA.

PrgmNr 3612 - A Translocation t(12;22) in Myxoid/Round Cell Liposarcoma

View session detail

Author Block: **V. Chia**[1], M. Zaki[2], W. Su[1], R. Tang[1], M. Guardiola[3], K. Eastwood[3], C. A. Tirado[4]; [1]UCLA, Los Angeles, CA, [2]Rancho Cucamonga, CA, [3]Baylor Scott & White, Temple, TX, [4]Baylor Scott & White Hlth.System- Temple, Temple, TX
Disclosure Block: **V. Chia:** None.
Myxoid/Round Cell Liposarcoma (MRCL) is characterized as a soft tissue sarcoma that is associated with unusual patterns of metastasis to extrapulmonary sites, such as bones and other soft tissue sites. Here, we present a case of a 48-year-old male patient, diagnosed with MRCL in March 2020. The patient presented with a grade 1 myxoid liposarcoma in his left leg. DNA FISH analysis showed variant rearrangements of the *EWSR1* (22q12) gene and loss of the 5â□□ *DDIT3* (*CHOP* 12q13) gene. The variant rearrangement showed one or two fusions with multiple separated (rearranged) signals. The *DDIT-EWSR1* rearrangement has been reported in MRCL. The variant rearrangements of the *EWSR1* (22q12) gene findings correlate with concurrent conventional cytogenetic findings and were described as nuc ish(EWSR1x2)(5'EWSR1 sep 3'EWSR1x1)[128/100],(5'EWSR1,3'EWSR1)x1~3(5'EWSR1 con 3'EWSR1x1~2)[57/100]. The variant rearrangements of the DDIT3 (CHOP 12q13) gene findings were described as nuc ish(5'DDIT3x1,3'DDIT3x2)(5'DDIT3 con 3'DDIT3x1)[195/200]. Molecular cytogenetic studies also showed a rearrangement of *EWSR1* (22q12) in 64% of nuclei and variant rearrangement in 31.5% of nuclei. A loss of *DDIT3*â□□s (12q13) 5â□□ signal was found in 97.5% interphase nuclei. Molecular pathology results indicated the patient was positive for *EWSR1* (exon 7) and *DDIT3* (exon 2) fusion. The patient underwent radiation therapy pre-resection of the myxoid liposarcoma. The most common form of MRCL is associated with t(12;16)(q13;p11), leading to *FUS-CHOP* and EWS-CHOP fusion proteins acting as aberrant transcription factors. The key element here is that this *DDIT-EWSR1* rearrangement led to a translocation t(12;22)(q13;q12) which is a rare cytogenetic event that led to the development of MRCL in this patient.

PrgmNr 3613 - Amplification of *CCND1* in Invasive High Grade Urothelial Carcinoma

View session detail

Author Block: **A. T. Reyes**[1], Y. Lin[1], M. Solanki[1], M. Guardiola[2], K. Eastwood[2], A. Cheng[1], C. A. Tirado[3]; [1]UCLA, Los Angeles, CA, [2]Baylor Scott & White, Temple, TX, [3]Baylor Scott & White Hlth.System-Temple, Temple, TX

Disclosure Block: **A.T. Reyes:** None.

Urothelial carcinomas (UC) are the most common form of bladder cancer, resulting from the transformation and excessive proliferation of urothelial cells. Herein, we report the case of ab 82 year old female patient diagnosed with high grade urothelial carcinoma (HGUC). She also presented a liver tumor. Tissue biopsy of a liver mass revealed a poorly-differentiated carcinoma with partial glandular differentiation and regions of a more solid component. Tumor cells were positive for CK7 and MOC31, while negative for CK20, GATA3, TTF1, CDX2, p63, PAX8, HepPar-1, Arginase-1, glypican, and mammaglobin, by immunohistochemical staining. DNA FISH analysis was performed on cells derived from a bladder tumor using the CCND1 (11q13) break apart probe, and the results were described as nuc ish(CCND1 amp)[50], indicating amplification but no rearrangement of *CCND1*. *CCND1* amplification is seen in ~10% of all bladder cancer patients and an independent risk factor of metastasizing cancer. Clinical correlation was indicated.

PrgmNr 3614 - Amplification of *NMYC* in Neuroblastoma

View session detail

Author Block: **S. Bottomley**[1,2], G. E. Yang*[1,2], K. Phan[1,2], A. Lozada[1,2], K. Eastwood[3], M. Guardiola[3], C. A. Tirado[3,2]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Intl. Circle of Genetics Studies, Los Angeles, CA, [3]Baylor Scott & White Hlth., Dept. of Pathology, Temple, TX
Disclosure Block: **S. Bottomley:** None.
We present a case study of a 20-month old female with a history of poorly differentiated neuroblastoma. Chromosomal analysis of the bone marrow revealed an abnormal karyotype including loss of chromosome 10, additional material of unknown origin on 2p, an interstitial deletion of 14q, a derivative chromosome 1 resulting from an unbalanced translocation involving 1p and 17q; and 50~150 double minute chromosomes in 11 of the 20 metaphase cells examined. This was characterized as 45,XX,der(1)t(1;17)(p13;q11.2),add(2)(p13),-10,del(14)(q24q31),50~150dmin[5]/44,sl,-X,add(3)(p21)[6]/46,XX[9]. DNA FISH analysis revealed amplification of the *MYCN* gene on 2p24.3 in all 50 of the examined nuclei. Increased expression of the *MYCN* oncogene is particularly common in neuroblastoma, as it characterizes approximately 20% of all cases. Expression often increases with tumor progression and is associated with an aggressive disease course and a poor prognosis. Despite advances in our knowledge of how the oncogene contributes to tumor development, it remains extremely difficult to treat directly. More research is necessary to identify feasible targets along the *MYCN* pathway to treat high risk neuroblastoma subtypes.

PrgmNr 3615 - An amplification of *EGFR (*7p12) in a patient with Glioblastoma

View session detail

Author Block: **E. Peng**[1], J. Yee[2], C. Tran[1], J. Glasser[3], K. Eastwood[4], M. Guardiola[4], C. A. Tirado[5]; [1]Univ. of California, Los Angeles, Fullerton, CA, [2]UCLA, IRVINE, CA, [3]The Intl. Circle of Genetic Studies, Fullerton, CA, [4]Baylor Scott & White, Temple, TX, [5]Baylor Scott & White Hlth.System- Temple, Temple, TX

Disclosure Block: **E. Peng:** None.

Glioblastoma Multiforme (GBM) is the most malignant and frequently occurring primary brain tumor out of all of the different types of primary astrocytomas. It presents with an extremely poor prognosis, with a median survival of 14 to 15 months from the diagnosis. Herein, we present an 83-year-old female patient with a right frontal brain mass. A craniotomy for the frontal brain mass was performed, which revealed a tumor with high-grade glioma, necrosis, atypia, and vascular proliferation. The patient was subsequently diagnosed with Glioblastoma Multiforme Grade IV (GBM). Molecular cytogenetic studies showed an amplification of the EGFR gene in 100% nuclei scored. These findings were described as: nuc ish(EGFR amp)[50/50],(MEGF6,TP73,ANGPTL1,ABL2,MAN2B1,ZNF44,GLTSCR1,GLTSCR2,CRX)x2[100]. Amplification of EGFR appears in 40-50% of individuals with Glioblastoma Multiforme Grade IV. Interestingly, although chromosomal deletions involving 1p36 and 19q13 are characteristic molecular features of solid tumors such as oligodendrocytes and mixed oligoastrocytomas, there was no evidence of a deletion of 1p36 or 9q13. Clinicopathologic correlation of these results was recommended.

PrgmNr 3616 - An automated classification system for copy number variants based on 2019 ACMG guidelines

View session detail

Author Block: **J. Ji**[1], R. Schmidt[1], W. Sherman[2], R. Peralta[2], M. Roytman[2], S. Shams[2], G. Raca[1]; [1]Children's Hosp. Los Angeles, Univ. of Southern California, Los Angeles, CA, [2]BioDiscovery Inc., El Segundo, CA

Disclosure Block: **J. Ji:** None.

Introduction: The ACMG, in collaboration with ClinGen, developed and published updated technical guidelines for interpretation and reporting of constitutional copy number variants. The updated guidelines were influenced by the ACMG guidelines for interpretation of sequence variants with a detailed algorithmic approach that looks at various sources of evidence in favor of a variant being pathogenic or benign with each evidence having a numerical weight that is added together to arrive at a final interpretation. Although the guideline is useful in arriving at a consistent classification, manually performing the review for every variant can be prohibitively time consuming. We have evaluated the utility of a software tool that automatically calculates the score for many of the evidence categories described by the guidelines to assess in improvement in review time and consistency in classification. The system provides the case reviewer flexibility to add additional evidence score(s) and modify the scores for each evidence category using professional expertise to arrive at a final interpretation. Methods: We compare both the quality and efficiency of using this system along with the guidelines versus manual classification without the guidelines. This analysis will be performed using 100 constitutional samples that had been previously reviewed and reported before the release of the guidelines. We compare the classification of these samples using the guidelines in the automated manner with manual adjustment and measure the rate of concordance and analyze the differences. We also measure the amount of time required to perform the interpretation using the semi-automated system following the guidelines to the manual approach without the guidelines and present the results. Results: Preliminary data demonstrated concordant results between automated and manual review for 9/10 cases. In addition, it took approximately 5 minutes to evaluate a case with the automated pre-classification system, which was generally faster than that with manual review alone. One case was classified as pathogenic manually but as a variant of uncertain clinical significance by automated system using the new ACMG guidelines requiring further investigation. Conclusions: Automated system for classification of copy number variants can facilitate systematic persistence of points assigned based on the evidence available for each variant. The system then tallies the points to help arrive at preliminary CNV classification. This automated system is quantitative, evidence-based, efficient, and may potentially minimize variability among the reviewers in the clinical diagnostic setting.

PrgmNr 3617 - Characterizing tumor-infiltrated immune cells with spatial context using RNAscope-immunohistochemistry co-detection workflow in FFPE tissues

View session detail

Author Block: **A. Dikshit**[1], J. Phatak[2], L. Hernandez[2], E. Doolittle[2], V. Murlidhar[2], B. Zhang[3], X-J. Ma[3]; [1]Bio-techne, Fremont, CA, [2]Bio-Techne, Newark, CA, [3]Bio-techne, Newark, CA
Disclosure Block: **A. Dikshit:** Salary/Employment; Bio-techne.
Complex tissues such as tumors are comprised of multiple cells types and extracellular matrix. Understanding the composition of immune cells in the tumor microenvironment (TME) can guide therapeutic intervention and predict treatment response. Thorough understanding of tissue dynamics and immune cell characterization requires a multi-omics approach. Simultaneous detection of RNA and protein using in situ hybridization (ISH) and immunohistochemistry/immunofluorescence (IHC/IF) can reveal cellular sources of secreted proteins, identify specific cell types, and visualize the spatial organization of cells within the tissue. However, a sequential workflow of ISH followed by IHC/IF frequently yields suboptimal protein detection because the protease digestion step in the ISH protocol. Here we demonstrate a newly developed integrated ISH/IHC workflow that can substantially improve RNA-protein co-detection to visualize tumor immune infiltrates at single-cell resolution with spatial context. To characterize tumor-infiltrating immune cells in a tumor TMA (tumor microarray), we utilized the RNAscope Multiplex Fluorescence assay in combination with the RNA-Protein Co-detection Kit to detect multiple immune cell populations. Immune cells such as macrophages, T cells and NK cells were detected using specific antibodies against CD68, CD8, CD4 and CD56, respectively. Precise characterization of these immune cells was achieved by using probes against targets such as *CCL5*, *IFNG*, *GNZB*, *IL-12*, *NCR1* etc. that not only help in identifying specific immune cells but also assist in determining their activation states. We identified subsets of T cells such as CD4+ regulatory T cells and CD8+ cytotoxic T lymphocytes. Additionally, we were able to determine the activation states of CD8+ T cells by visualizing the expression of *IFNG* and *GZMB.* Furthermore, infiltrating macrophages were identified by detecting the CD68 protein expression while the M1 and M2 subsets were differentiated by detecting the M2-specific target RNA for *CD163.* Similarly, NK cells were identified by detecting CD56 protein in combination with *CCL5* and *NCR1* RNA expression. Interestingly, the degree of infiltration of the different immune cell populations varied based on the tumor type. In conclusion, the new RNAscope-ISH-IHC co-detection workflow and reagents enable optimized simultaneous visualization of RNA and protein targets by enhancing the compatibility of antibodies - including many previously incompatible antibodies - with RNAscope . This new workflow provides a powerful new approach to identifying and characterizing tumor infiltrating populations of immune cells.

PrgmNr 3618 - Monosomy 21, a rare finding in an elder patient with non-CLL-like monoclonal B-cell lymphocytosis

View session detail

Author Block: **W. Yeh**[1], D. Mrugala[2], L. Nabipur[1], H. Robinson[1], A. Zhao[1], K. Eastwood[3], M. Guardiola[3], C. A. Tirado[4]; [1]UCLA, Los Angeles, CA, [2]Texas A&M Coll. of Med., Houston, TX, [3]Baylor Scott & White, Temple, TX, [4]Baylor Scott & White Hlth.System, Temple, TX

Disclosure Block: **W. Yeh:** None.

Monoclonal B-cell lymphocytosis (MBL) can be classified as either CLL-like MBL, atypical MBL, or non-CLL-like MBL. Non-CLL-like MBL, which makes up less than 20% of MBL cases, is rare and not well characterized. However, an association with splenic marginal zone lymphoma (SMZL) has been suggested. Aberrations such as isochromosome 17q resulting in the loss of p53 and 7q abnormalities have previously been reported. Here we report a case of an 85-year-old man presenting with non-CLL-like MBL. Immunophenotyping revealed a small lambda-restricted B lymphoid population which tested positive for CD19 and CD20, while CD5 tested negative. Patient karyotype showed the presence of monosomy 21 unaccompanied by additional chromosomal changes in 3 of the 35 cells examined. Monosomy 21 as a sole abnormality has rarely been documented in non-CLL-like MBL, which already lacks established cytogenetic features. The significance of this finding remains unknown.

PrgmNr 3620 - Outcomes from integrating an accessible delivery model for hereditary cancer risk assessment and genetic testing in populations with barriers to access

View session detail

Author Block: **L. Amendola**[1], E. Shuster[2], M. C. Leo[2], M. O. Dorschner[3], B. A. Rolf[1], B. H. Shirts[4], M. J. Gilmore[2], J. Zepp[2], T. L. Kauffman[5], K. F. Mittendorf[5], S. Okuyama[6], C. A. Bellcross[7], C. L. Jenkins[2], G. Joseph[8], L. Riddle[9], S. Syngal[10], C. Ukaegbu[10], K. A. Goddard[2], B. S. Wilfond[11], G. P. Jarvik[12]; [1]Univ. of Washington, Seattle, WA, [2]Kaiser Permanente Northwest, Portland, OR, [3]Michael Dorschner, Seattle, WA, [4]Univ of Washington, Seattle, WA, [5]Kaiser Permanente, Portland, OR, [6]Denver Hlth.and Hosp. Authority, Denver, CO, [7]Emory Univ Sch. of Med., Atlanta, GA, [8]Univ California San Francisco, San Francisco, CA, [9]Univ. of California San Francisco, San Francisco, CA, [10]Dana Farber Cancer Inst., Boston, MA, [11]Seattle Children's Res. Inst., Seattle, WA, [12]Univ Washington Med Ctr., Seattle, WA
Disclosure Block: **L. Amendola:** None.
Access to cancer genetic testing and services is inequitable. The CHARM study evaluated a multimodal intervention to address healthcare disparities in cancer genetics. Patients from clinics that serve populations with access barriers, who screened at risk for a hereditary cancer syndrome via online, literacy- and culturally-adapted cancer history collection tools (B-RST™ 3.0, PREMM$_5$™), or with limited family history information, were offered exome-based panel testing for cancer risk and medically actionable secondary findings. We used descriptive statistics, electronic health record review, and inferential statistics to explore participant characteristics and findings, healthcare consultations and actions related to P/LP variants disclosed, and variables predicting category of findings, respectively. Of 967 participants consented for genetic testing and sent a saliva collection kit, 841 (87%) returned their kit to the lab. Overall, 40 (5%) participants had a P/LP cancer risk variant, including 10 in *BRCA1/2* and 7 in Lynch syndrome genes; 9 (1%) had a medically actionable secondary finding. For 11/34 (32%) participants disclosed a P/LP cancer risk finding, the variant was previously known in the participant (N=7) and/or their family (N=4). Of 42 participants disclosed any P/LP variant, 16 (38%) were returned after March 1st 2020, when healthcare access limitations due to COVID-19 began. Recommended consultations with medical genetics and other providers were completed for 9/32 and 14/15 participants, respectively, after an average of 17 months post disclosure. Recommended actions (ex. breast MRI) were completed by 17/25 participants. Participant history of breast, ovarian, colon and/or endometrial cancer and score on the PREMM$_5$™ tool were each associated with category of findings (history and colon cancer risk CramÃ©r's V=.11, Fisher's exact=.02; history and breast cancer risk CramÃ©r's V=.13, Fisher's exact=.01; PREMM$_5$™ and colon cancer risk CramÃ©r's V=0.22, Fisher's exact<.001 the charm study approach provided an accessible delivery model for hereditary cancer risk assessment and genetic testing. rate of p variants in genes identified this population informs expectations guidelines results disclosure overlapped with covid-19 related care restrictions however interventions to increase adherence finding recommendations are likely still necessary. adapting provider-facing family history collection tools into patient-facing electronic applications can help identify patients at increased cancer.>

PrgmNr 3621 - Patients with Birt-Hogg-DubÃ© syndrome and papillary renal cell carcinoma reveal a rare intragenic deletion in *FLCN*

View session detail

Author Block: **M. Evans**[1], A. Sisk[2], S. Liu[2], B. Shuch[2], F. Quintero-Rivera[3]; [1]The Univ. of Texas M.D. Anderson Cancer Ctr., Houston, TX, [2]Univ. of California, Los Angeles, CA, [3]Univ. of California, Irvine, CA
Disclosure Block: **M. Evans:** None.
Birt-Hogg-DubÃ© syndrome (BHDS) is an inherited, autosomal dominant, multisystem disorder involving the *FLCN* gene on chromosome 17p11.2. *FLCN* encodes a tumor suppressor protein that regulates the mTOR pathway, and genetic alterations lead to multiple cutaneous, pulmonary, and renal findings with high penetrance and variable expressivity. We describe two unrelated patients with BHDS who developed a renal cell carcinoma (RCC) papillary subtype which is not typically associated with this syndrome, leading to detection of a rare *FLCN* intragenic deletion. A 52-year-old man and a 78-year-old man presented with abdominal and flank pain that they had experienced for up to one year. Both individuals were shown to have multiple pulmonary cysts on imaging, as well as flesh-colored to white papules on the face and neck that were consistent with fibrofolliculomas on biopsy. Moreover, the patients were found to have unilateral kidney tumors, with the younger individual having evidence of lymphadenopathy, inferior vena cava invasion, and pulmonary metastasis. Both patients had surgical resection demonstrating papillary RCC. The overall constellation of findings in these individuals was highly suggestive of BHDS, and genetic testing revealed a p.Lys508del (c.1522_1524delAAG) in-frame deletion in *FLCN* (NM_144997.5). The detected alteration in *FLCN* has been previously reported in only one family of patients with a BHDS phenotype. A functional study has demonstrated that the *FLCN* in-frame deletion p.Lys508del results in a truncated folliculin protein with significantly decreased stability. Moreover, among the more than 55 tumors presented in a retrospective survey of BHDS patients, only one case of papillary RCC was documented. Information regarding the prognosis or behavior of this tumor subtype in BHDS is unavailable; unfortunately, the younger of our two patients passed away from his cancer. Whether p.Lys508del in BHDS is associated with the papillary variant of RCC remains unknown and warrants further investigation.

PrgmNr 3622 - Rapid whole genome sequencing (rWGS) on a novel digital microfluidic system: â□□on demandâ□□ personal automation for diagnosis of critically ill infants

View session detail

Author Block: **Y. Ding**[1], K. Cunningham[2], J. Yoo[2], S. Batalov[1], A. Johnson[1], E. Cervantes[2], A. Barner[2], A. E. Carvajal[2], L. Cheng-Chang[2], S. Margeridon[2], M. Jebrail[2], W. Benson[1], F. Christodoulou[2], S. Kingsmore[1]; [1]Rady Children's Inst. for Genomic Med., San Diego, CA, [2]Miroculus, San Francisco, CA
Disclosure Block: **Y. Ding:** None.
Genetic disorders and congenital anomalies are a leading cause of morbidity and mortality in infants, affecting ~2% of live births in the United States. There is now strong evidence that rapid, clinical Whole Genome Sequencing (rWGS) at presentation of critically ill infants enables precision medicine interventions for many of these genetic diseases that improves outcomes and reduces cost of care. There is now a need to scale rWGS for national adoption. However, manual rWGS library preparation is laborious and lengthy that cannot be delayed for sample batching, and that could benefit from a system that consistently generates high quality DNA libraries for rWGS with minimal manual effort. Here, we report a novel solution: digital microfluidic (DMF)-based system, Miro Canvas. Inter-sample contamination is prevented by constructing libraries within a single-use, electronics-free cartridge. Miro Canvas utilizes electromechanical forces to move, merge, mix and dispense microliter-volume fluids across a surface of patterned electrodes in an automated fashion. We automated 2 PCR-free rWGS library preparation protocols using Miro Canvas (Miro PCR-free WGS libraries with mechanical fragmentation and Illumina DNA PCR-free Prep Kit with tagmentation) and compared quality with standard manual and high-throughput liquid handler-based methods. The input was 500 ng purified genomic DNA samples from reference materials (Coriell Repository NA12878) or patient samples with research consent at the Rady Childrenâ□□s Institute for Genomic Medicine. Library quality was assessed by qPCR or ssDNA Qubit and WGS was performed using NovaSeq6000 systems. Library yield, insert size, and gene coverage was comparable to that of standard methods. In conclusion, a DMF-based system generated high-quality rWGS libraries with minimal manual effort. For singleton, duo, and trio samples from critically ill children, this system provides a novel solution for automated library preparation when waiting for centralized or batched sample processing is not a viable option.

PrgmNr 3623 - Simplified and robust library construction for high-throughput HiFi sequencing for human variant detection

View session detail

Author Block: **H. Dhillon**; Pacific BioSci.s, Menlo Park, CA
Disclosure Block: **H. Dhillon:** Salary/Employment; Pacific Biosciences.
HiFi sequencing is proving to increase understanding of undiagnosed groups in rare and inherited disease research. Laboratories worldwide are adopting HiFi reads due to high accuracy (>Q30) and long read lengths (up to 25 kb) making it ideal for the accurate detection of single nucleotide variants and large structural variants. However, the quantity of DNA available for a cohort can place constraints on a given study especially when the source of the sample is already deceased or access to additional DNA is not possible. In this study, we will address this common problem by providing an alternate workflow such that DNA quantities as low as 0.9 Âµg can be sequenced to generate >10-fold coverage of HiFi data.Currently, standard library preparation protocols for successful HiFi sequencing require 5 Âµg of high-quality, high-molecular weight human genomic DNA (gDNA). Size-selection using traditional gel-based system is recommended to achieve highly accurate long reads with subread lengths between 15 kb -18 kb. Using this approach, post-size selection recovery yield is often low (10%) making it impossible to generate sufficient library from 24 GB of HiFi reads per SMRT Cell 8M run on the Sequel II System. The mean HiFi read lengths were ~16 kb with a median read quality of Q35. Furthermore, a bead-based method enables improvements to automation so that HiFi library construction steps from shearing DNA to size-selection can be fully automated enabling the preparation of up to 96 samples in parallel. In summary, we have developed a simple and fast workflow for generating high-quality HiFi reads on the Sequel II System for human variant detection using challenging real-world samples with limited DNA amount.

PrgmNr 3624 - Utilizing Somatic Tumor Profiling To Identify Candidate Patients For Germline Genetic Testing In A Community Setting Reveals Barriers To Referral To Genetic Counseling

View session detail

Author Block: **S. Darabi**[1], J. Homer[1], C. Weatherill[1], V. Dalili-Shoaie[1], M. Demeure[1,2], D. Braxton[1]; [1]Hoag Family Cancer Inst., Newport Beach, CA, [2]Translational Genomics Res. Inst., Phoenix, AZ
Disclosure Block: **S. Darabi:** Consultant/Consulting Fees/Other Remuneration; oncolense, Bayer.
Background: Between 5 and 10 percent of patients with cancer have an inherited predisposition to cancer. Some patients who harbor pathogenic or likely pathogenic germline variants would not meet current clinical testing guidelines for genetic testing. The use of somatic mutational profiling has increased and has become the standard of care for many cancers, but most somatic profiling tests do not include paired germline testing. The presence of a suspected germline mutation in a cancer predisposition gene requires confirmatory germline testing. We screened somatic molecular profiling results for suspected germline genetic variants in an effort to increase the detection of inherited cancer predisposition gene mutations. **Methods:** We classified potential germline mutations into 3 tiers based on the likelihood of pathogenicity. Per our routine practice of medicine, all somatic sequencing panel results were reviewed based on the IRB-approved protocol for our designated Tier one genes (*ATM, BAP1, BMPR1A, BRCA1, BRCA2, BRIP1, MSH2, MSH6, MUTYH, MYH11, PALB2, RUNX1, SDHAF2, SDHB, SDHC,* and known founder mutations in additional genes). Treating physicians for eligible patients were contacted, and if they agreed, then the patients were offered participation in this study. Enrolled patients had genetic counseling and genetic testing for a multi-gene panel at no cost to the patient. **Results:** We screened 1,781 cases, and 167 patients were eligible for the study. In all, 13 patients were enrolled, but only 11 completed testing. Of these, 6 had germline mutations (54.5%) confirmed, 3 had a variant of uncertain significance (27.3%), and 2 had negative testing (18.2%). Some patients were not referred by their treating physicians based on their own assessment. Others were hesitant to enroll due to concern about the privacy of genetic information. Other confounding factors may affect our accrual, including the COVID-19 pandemic. **Conclusion:** Implementation of a somatic profiling-based screening program to identify inherited cancer predisposition syndromes is feasible; however, there are barriers to genetic counseling and testing. Our study showed only 6.6% of eligible patients underwent genetic counseling and testing even when funding removed cost as a barrier. Other barriers need to be overcome to increase the utilization of genetic counseling.

PrgmNr 3625 - Verification of an *EZH2* mutation test from formalin-fixed paraffin-embedded (FFPE) follicular lymphoma tumor tissue specimens

View session detail

Author Block: **S. Rosenthal**, A. Le, J. Yo, C. Ma, D. Sugganth, F. Lacbawan; Quest Diagnostics, San Juan Capistrano, CA
Disclosure Block: **S. Rosenthal:** None.
Background: Follicular lymphoma (FL) is the second most common non-Hodgkin lymphoma among adults in developed countries. *EZH2* mutations occur in more than 25% of FL patients, and different types of EZH2 methyltransferase inhibitors have been developed. Screening *EZH2* mutations to help select patients suitable for EZH2-target therapy is important in personalized precision therapy. This study evaluates the performance characteristics of an FDA-approved *EZH2* mutation test in a clinical setting to aid management of patients with FL eligible for treatment with tazemetostat, an inhibitor of EZH2. **Methods**: A total of 13 unique positive and 8 unique negative samples were analyzed for a total of 43 and 34 trials, respectively, by the Roche cobas *EZH2* mutation test. The cobas *EZH2* test is an FDA-approved, real-time allele-specific PCR test for qualitative detection of single-nucleotide mutations for *EZH2* (Y646N/F/H/S/C, A682G, and A692V) in DNA extracted from formalin-fixed paraffin-embedded (FFPE) human FL tumor tissue specimens. Sample pass rate, inter- and intra-assay precision, and accuracy were analyzed. For the accuracy study, results were compared to the expected mutation calls by a validated next-generation sequencing (NGS) method. **Results**: All 77 samples from 21 unique FFPE specimens were successfully analyzed by 2 different scientists in 4 independent set-ups. Variant calls were 100% concordant among intra- and inter-assay replicates for all positive and negative samples. Positive percent agreement was 100% (95% CI: 91.8%-100%), and negative percent agreement was 100% (95% CI: 90.0%-100%) in comparison to NGS result. **Conclusion**: The cobas *EZH2* mutation test demonstrated a robust and reliable assay performance for the detection of the targeted *EZH2* mutations in patients with FL in a clinical setting.

PrgmNr 3626 - A large genome-wide association study for melanoma-specific survival reveals two novel loci linked to *PSEN2,* and *ADCK3* genes

View session detail

Author Block: **M. Seviiri**[1,2], R. A. Scolyer[3,4,5], J. F. Thompson[3,5,6], G. V. Long[3,7,5], S. Macgregor[8,2], M. H. Law[8,2]; [1]Statistical Genetics, QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]Sch. of BioMed. Sci., Faculty of Hlth., and Ctr. for Genomics and Personalised Hlth., Queensland Univ. of Technology, Brisbane, Australia, [3]Melanoma Inst. Australia, The Univ. of Sydney, Sydney, Australia, [4]Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hosp. and NSW Hlth.Pathology, Australia, Sydney, Australia, [5]Faculty of Med. and Hlth., The Univ. of Sydney, Sydney, Australia, [6]Dept. of Melanoma and Surgical Oncology, Royal Prince Alfred Hosp., Camperdown, Australia, [7]Dept. of Med. Oncology, Royal North Shore Hosp., Sydney, Australia, [8]QIMR Berghofer Med. Res. Inst., Brisbane, Australia

Disclosure Block: **M. Seviiri:** None.

Background: Melanoma is the deadliest skin cancer. Important prognostic factors include age, site of the primary tumour, primary tumour thickness, primary tumour ulceration, mitotic rate, and the stage of the tumour. The role or relative importance of germline genetic variation in melanoma survival is poorly understood. Objective: To identify germline genetic variants that influence melanoma-specific survival (MSS). Methods: First, we conducted two genome-wide association studies (GWAS) of melanoma-specific survival (MSS) in the Melanoma Institute Australia cohort (MIA; 5,762 patients with melanoma and 800 melanoma-specific deaths) and in the UK Biobank (5,220 melanoma patients and 241 melanoma-specific deaths) cohort. Using Cox proportional-hazard modelling, hazard ratios (HR) were computed adjusting for age, sex and the first ten ancestral principal components. Next, we conducted a GWAS-meta-analysis (N=10,982 including 1,041 melanoma-specific deaths) using a fixed-effects inverse-variance weighted model. Results: Two independent genome-wide significant (P-8) loci for MSS with lead SNPs rs41309643/G (HR=2.09, 95% CI = 1.61-2.71, P=2.08x10^{-8}, EAF=2%) on chromosome 1 (1q42.13) and rs75682113/C (HR=2.38, 95% CI = 1.77-3.21, P=1.07x10^{-8}, EAF=2%) on chr 7 (7p14.1). rs41309643/C is an intron of the *PSEN2* gene; rare mutations in this gene influence Alzheimer's disease susceptibility, and cardiomyopathy dilated type 1V risk. eQTL analysis revealed another gene *ADCK3* that is induced by *p53* in response to DNA damage; inhibition of *ADCK3* counteracts p53-induced apoptosis. The other novel variant rs75682113 is in an intron of the *SUGCT* gene; mutations in this gene have been associated with glutaric aciduria type 3 disease susceptibility. Conclusion: A large GWAS analysing MSS revealed novel loci; while needing functional confirmation, we identified potential candidate genes underlying these loci whose biology is linked to tumour progression, as well as premature deaths due to congestive heart failure.

PrgmNr 3627 - A segregation analysis of 17,425 population-based breast cancer families: Implications for breast cancer genetic susceptibility and risk prediction

View session detail

Author Block: **S. Li**[1], A. Lee[2], R. J. MacInnis[3], L. Dorling[2], S. Carvalho[2], T. Nguyen-Dumont[4,1], G. S. Dite[1], G. G. Giles[3], P. Pharoah[5], M. C. Southey[6], D. F. Easton[2], J. L. Hopper[7], A. C. Antoniou[2]; [1]The Univ. of Melbourne, Melbourne, Australia, [2]Univ. of Cambridge, Cambridge, United Kingdom, [3]Cancer Council Victoria, Melbourne, Australia, [4]Monash Univ., Clayton, Australia, [5]Univ Cambridge, Cambridge, United Kingdom, [6]Monash Univ., Melbourne, Australia, [7]Univ Melbourne, Carlton, Australia
Disclosure Block: **S. Li:** None.
Background Breast cancer risk models used in family cancer clinics rely on accurate modelling of the familial relative risks; however, rare pathogenic variants in known breast cancer susceptibility genes, together with known common genetic variants, do not fully explain the familial aggregation of breast cancer. We aimed to investigate plausible genetic models for the residual familial aggregation.
Methods We used data on 17,425 UK and Australian families ascertained through population-based sampling of women diagnosed with breast cancer (probands). 86% of probands were screened for pathogenic variants in *BRCA1*, *BRCA2*, *PALB2*, *CHEK2*, *ATM* and *TP53* using gene-panel sequencing. We conducted complex segregation analyses and fitted genetic models in which breast cancer incidence depended on the effects of pathogenic variants in the known susceptibility genes, other unidentified major genes and a normally distributed polygenic component. Maximum likelihood estimation was used to estimate the allele frequencies and breast cancer risk associated with pathogenic variants in the genes. We computed the age-specific familial relative risks (FRRs) predicted by the genetic models, defined as the ratio of the risk of a woman with an affected 1st-degree relative to the population risk.
Results 881 probands carried pathogenic variants in the six known genes. Across ages, on average, pathogenic variants in *BRCA1* and *BRCA2* explained 16% and pathogenic variants in *PALB2*, *CHEK2*, *ATM* and *TP53* explained 5% of the familial aggregation of breast cancer. After allowing for these variants, the best fitting model for the residual aggregation involved a recessively inherited allele with frequency of 13% (95% CI: 0.6-21%) and penetrance of 69% (95% CI: 43-91%) by age 80 for homozygous carriers, explaining 18% of the residual familial aggregation, and a polygenic component with an age-independent variance of 1.27 (95% CI: 0.96-1.63). A risk model based on the best fitting model predicted FRRs to be 3.4, 2.0, 1.7, 1.5 and 1.4 for a woman aged in her 30s, 40s, 50s, 60s and 70s with her mother affected at the same age, and the corresponding FRRs associated with an affected sister were 3.7, 2.3, 1.8, 1.6 and 1.4. The predicted FRRs were consistent with the FRRs estimated by epidemiological studies of the Collaborative Group in Hormonal Factors in Breast Cancer.
Conclusion In addition to the pathogenic variants in known breast cancer susceptibility genes, and polygenic risk scores, unidentified major usceptibility genes might exist which explain breast cancer familial aggregation. Our findings have implications for attempts to identify new breast cancer susceptibility genes and risk prediction.

PrgmNr 3628 - Associations between tumor inflammation, survival, and somatic mutations in colorectal cancer patients

View session detail

Author Block: **H. Yin**[1,2], T. A. Harrison[2], S. S. Thomas[2], C. L. Sather[3], A. L. Koehne[4], R. C. Malen[2], A. M. Reedy[2], M. A. Wurscher[2], L. Hsu[5], A. I. Phipps[2], S. H. E. Zaidi[6], P. A. Newcomb[2], U. Peters[2], J. R. Huyghe[2]; [1]Inst. for Publ. Hlth.Genetics, Univ. of Washington, Seattle, WA, [2]Publ. Hlth.Sci. Div., Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [3]Genomics & Bioinformatics, Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [4]Experimental Histopathology Shared Resource, Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [5]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [6]Ontario Inst. for Cancer Res., Toronto, ON, Canada
Disclosure Block: **H. Yin:** None.

Colorectal cancer (CRC) is the third most common cancer and a leading cause of cancer-related death in the United States. The anti-tumor immune response plays a key role in tumor progression and survival. The T cell-inflamed gene expression profile (GEP) is a biomarker predicting response to checkpoint inhibitor immunotherapy. We evaluated its ability in predicting disease-specific death and associations with somatic mutations among sporadic CRC patients. This study included 84 incident CRC cases from the Seattle Colon Cancer Family Registry with formalin-fixed paraffin-embedded (FFPE) tumor tissues collected between 1997-2003. Expression profiling of 770 genes was performed using Nanostringâ□□s nCounter PanCancer IO 360 panel (NanoString Technologies, Seattle, WA). The T cell-inflamed GEP was calculated as a weighted average of housekeeping gene-normalized expression counts of 18 genes (Cristescu et al., 2018 Science). Somatic mutations were identified by targeted DNA sequencing using a custom AmpliSeq panel (Zaidi et al., 2020 Nat Commun). The Cox proportional hazard model was used for survival analyses and linear regression was used for somatic mutation association analyses. The T cell-inflamed GEP was positively associated with log-transformed tumor mutation burden (r = 0.37, P = 0.0008), hypermutation status (P = 0.001) and microsatellite instability-high status (MSI-H) ($P = 2.67 \times 10^{-6}$). Higher T cell-inflamed GEP had favorable CRC-specific survival (hazard ratio (HR) per SD unit = 0.50, 95% CIs = 0.31 - 0.80, P = 0.004) after adjusting for age at diagnosis, sex and hypermutation and/or MSI status. The result remained significant among the non-hypermutated and microsatellite stable (MSS) cases (HR = 0.53, 95% CIs = 0.33 - 0.85, P = 0.008). Analysis of recurrently mutated genes having at least 10 mutation carriers across cases, suggested T cell-inflamed GEP is positively associated with *RYR1* (16 carriers, P = 0.0025, false discovery rate (FDR) adjusted P = 0.028), and negatively associated with *APC* (44 carriers, P = 0.0097, FDR adjusted P = 0.053). However, after adjusting for hypermutation and/or MSI-H status, which act as confounders, the two associations were no longer significant. Our results show that the T cell-inflamed GEP predicts CRC prognosis independent of hypermutation and MSI status. This provides further evidence that the T cell-inflamed GEP is a measure of T cell activation in the T cell-inflamed tumor microenvironment that is distinct from tumor antigenicity. At the time of writing, we are finalizing a differential expression analysis that aims to identify genes and pathways that may modulate T cell activation and tumor immune evasion.

PrgmNr 3629 - Associations of genetically predicted blood protein biomarkers with prostate cancer risk: a study using comprehensive protein genetic prediction modes

View session detail

Author Block: **L. Wu**[1], D. H. Ghoneim[1], J. Zhu[1], X. Xu[1], Y. Sun[1], P. Surendran[2], T. Liu[3], S. Fahle[2], A. S. Butterworth[2], C. Wu[4]; [1]Univ. of Hawaii Cancer Ctr., HONOLULU, HI, [2]Univ. of Cambridge, Cambridge, United Kingdom, [3]Pacific Northwest Natl. Lab., Richland, WA, [4]Florida State Univ., Tallahassee, FL
Disclosure Block: **L. Wu:** None.
Prostate cancer (PCa) is the second most frequently diagnosed malignancy among males. Identification of biomarkers is critical for understanding the pathogenesis of this common cancer. Several blood protein markers have been linked to PCa risk in previous studies, but these studies have assessed only a small number of protein biomarkers usually in small set(s) of samples. To identify novel circulating protein biomarkers of PCa, we performed a large study in 79,194 prostate cancer cases and 61,112 controls of European ancestry included in PRACTICAL/ELLIPSE consortia by using comprehensive protein genetic prediction models as instruments. Leveraging genome and plasma proteome data of 2,841 healthy European descendants included in the INTERVAL study, we established models using four methods (best linear unbiased predictor, elastic net, LASSO, and top1) to predict protein levels based on genetic variants. For each protein, of the four sets of models developed, the model showing the highest prediction performance (R^2) was retained. We selected 2,157 built protein models with a prediction performance (R^2) of >0.01 for association analyses with PCa risk. We observed associations between predicted concentrations of 45 proteins and PCa risk at a false discovery rate of P-value-5). Twenty-one of the identified proteins showed positive associations and 24 showed inverse associations (*P*-values from 1.03 Ã 10^{-3} to 5.92 Ã 10^{-152}). Besides 24 proteins that were previously reported in our earlier study using protein quantitative trait loci as instruments, we identified 21 novel proteins in this study. Pathway enrichment analysis showed that genes encoding these proteins were significantly enriched in cancer related pathways, including STAT3 Pathway, CDK5 Signaling, p38 MAPK Signaling, and GP6 Signaling Pathway. In conclusion, we identified 45 protein biomarker candidates for PCa risk with the potential to improve our understanding of the etiology of PCa.

PrgmNr 3630 - Breast cancer polygenic risk scores and rare variants in Latinas

View session detail

Author Block: **J. L. Nierenberg**[1], A. Adamson[2], Y. C. Ding[2], Y. Shieh[1], D. Hu[1], S. Huntsman[1], E. M. John[3], G. Torres-Mejia[4], C. A. Haiman[5], L. H. Kushi[6], C. N. Ricker[7], L. Steele[2], R. Lee[1], J. N. Weitzel[8], L. Fejerman[9], S. L. Neuhausen[2], E. Ziv[1]; [1]Univ. of California San Francisco, San Francisco, CA, [2]Beckman Res. Inst. of City of Hope, Duarte, CA, [3]Stanford Univ. Sch. of Med., Stanford, CA, [4]Inst. Natl. de Salud Pul□blica, Cuernavaca, Mexico, [5]Univ. of Southern California, Los Angeles, CA, [6]Kaiser Permanente Northern California, Oakland, CA, [7]USC Norris Comprehensive Cancer Ctr., Los Angeles, CA, [8]Sierra Madre, CA, [9]Univ. of California Davis, Davis, CA

Disclosure Block: **J.L. Nierenberg:** None.

Introduction: Polygenic risk scores (PRS), assembled from common single nucleotide polymorphisms (SNPs), can be used to predict breast cancer risk. Individually, pathogenic variants (PVs) in high and intermediate penetrance breast cancer susceptibility genes are rare but have large effects on disease risk. Few studies have examined PRS and PVs in susceptibility genes together, and most previous breast cancer genetics studies have been conducted in European ancestry populations. Here, we report findings on the combined effects of PRS and PVs in Latinas.

Methods: We conducted a pooled case-control analysis of breast cancer in Latinas from the San Francisco Bay Area, Los Angeles, and Mexico (1,776 cases and 1,589 controls). Case ascertainment included 432 participants from high-risk studies (age below 50, family history, or bilateral breast cancer) and 1,344 from general population studies. We assembled a 180-SNP PRS from known breast cancer SNPs (P-8). We determined presence of a rare PV in 9 known breast cancer risk genes (*ATM*, *BARD1*, *BRCA1*, *BRCA2*, *CHEK2*, *PALB2*, *PTEN*, *RAD51C*, and *TP53*). We used multivariable logistic regression and area under the receiver operating characteristic curve (AUC) to examine the relationships between PRS, PV status, and breast cancer risk. All analyses were adjusted for age, study, and ancestry. Secondary analyses were stratified by age, family history, or indigenous ancestry above or below the median.

Results: Higher PRS was associated with higher risk of breast cancer, with an odds ratio of 1.6 (95% confidence interval [CI]: 1.5-1.7) per PRS standard deviation and an AUC of 0.61 (95% CI: 0.60-0.63). PVs were found in 125 case and 22 control participants. Having a PV was associated with a 5.9-fold (95% CI: 3.8-9.6) increased odds of breast cancer. The AUC for breast cancer improved to 0.64 (95% CI: 0.62-0.66) when PVs were added to the PRS model. Results were similar among 255 cases and 129 controls with family history of breast cancer and among those with indigenous ancestry above and below the median. Among 599 cases and 387 controls under age 50 years, the AUC for the PRS-rare variant model increased to 0.69 (95% CI: 0.65-0.72). Among all cases, those with a PRS below the median had a 1.9-fold (95% CI: 1.4-2.7) increased odds of having a PV.

Conclusion: We found that adding PV status to a PRS model improves prediction of breast cancer, indicating that there may be clinical utility in examining PRS and PVs together, especially among young women. The lower PRS among carriers of a PV is likely an effect of case ascertainment and could be useful in epidemiological design of new gene discovery efforts.

PrgmNr 3631 - Genetic risk of second primary cancer in breast cancer survivors: the Multiethnic Cohort Study

View session detail

Author Block: **F. Chen**[1], A. J. de Smith[2], S. L. Park[3], L. R. Wilkens[3], C. A. Haiman[4]; [1]Ctr. for Genetic Epidemiology, Dept. of Preventive Med., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [2]Univ. of Southern California, Los Angeles, CA, [3]Cancer Epidemiology Program, Univ. of Hawaii, Honolulu, HI, [4]Ctr. for Genetic Epidemiology, Los Angeles, CA
Disclosure Block: **F. Chen:** None.
Women who have had breast cancer in the past are at increased risk of developing second primary cancer (SPC), in particular second primary breast cancer (SPBC). Identifying risk factors for SPCs is essential for cancer prevention efforts, especially with an increasing population of breast cancer survivors. Pathogenic variants (PVs) in *BRCA1* or *BRCA2* genes are known to contribute to the excess risk of SPBC among female breast cancer survivors, but little is known regarding the genetic risk of developing any SPC, and whether this varies across populations. In a sub-cohort of 3,223 Multiethnic Cohort Study (MEC) female participants, from five racial/ethnic populations (White, African American, Japanese American, Latino and Native Hawaiian) who had incident invasive breast cancer during follow-up, we conducted a prospective analysis to investigate the associations of genetic variants with risk of SPC. Sequencing of 37 cancer predisposition genes was carried out by the CAnceR RIsk Estimates Related to Susceptibility (CARRIERS) study. Breast cancer case status was identified by linkage to the SEER registry and SPC status was determined by SEER multiple primary guidelines. Out of 3,223 MEC primary breast cancer cases, 719 (22.3%) women later developed a SPC, of which 323 (10.0%) cases were SPBC. The Cox proportional hazard model was used to assess risk. We found that PVs of four genes were significantly enriched in women with SPC: *BRCA1* (hazard ratio, HR=2.3 [95% CI:1.1-4.7], P=0.022), *CHEK2* (HR=3.1 [1.5-6.7], P=0.003), *ERCC2* (HR=3.6 [1.3-9.7], P=0.012), and *PPM1D* (HR=3.4 [1.7-6.6], $P=3.4x10^{-4}$) after adjusting for potential confounders. Women carrying a PV in any of these significant genes on average had a 3.0-fold increased risk of SPC (95% CI=2.1-4.4, $P=2.0x10^{-8}$) compared to non-carriers, and this association was consistently observed across ethnic groups. In the analysis by SPC subtypes, the significant association of *ERCC2* gene was more profound in women with SPBC (HR=5.8 [1.8-18.7], P=0.003), whereas the significant enrichment of *CHEK2* PVs was specific to women with non-breast SPC (HR=5.4 [2.4-12.4], $P=6.1x10^{-5}$). There was no evidence of effect modification by age at first breast cancer diagnosis, hormone receptor status or treatment (chemotherapy, radiotherapy, or hormonal therapy). Our multiethnic study identified four genes associated with the development of SPC in female breast cancer survivors, of which *ERCC2* and *PPM1D* have not previously been implicated in SPC or SPBC. Future studies are warranted to validate our results. These findings suggest that closer monitoring of SPC may be needed for women carrying PVs in these genes.

PrgmNr 3632 - Improving risk prediction: Validation and exploration of breast cancer PRSs in a Midwest-American population

View session detail

Author Block: **J. Slunecka**[1,2], J-J. Hottenga[3], J. Beck[1], B. Johnson[4], A. Van Asselt[1], H. Nickles[5], C. Ageton[5], C. Finnicum[1], A. Krie[6], D. Boomsma[3], E. A. Ehli[7]; [1]Avera Inst. for Human Genetics, Sioux Falls, SD, [2]Univ. of South Dakota Sanford Sch. of Med., Sioux Falls, SD, [3]Vrije Univ.it Amsterdam, Amsterdam, Netherlands, [4]Avera Inst. for Human Genetics, Sioux Falls, SD, [5]Avera McKennan Cancer Res., Sioux Falls, SD, [6]Avera McKennan Hosp., Sioux Falls, SD, [7]Avera Mckennan Hosp., Sioux Falls, SD
Disclosure Block: **J. Slunecka:** None.
INTRODUCTION: Breast cancer (BC) is the most common cancer among women and is classified as a complex disease. Advances in population genomics have led to the development of polygenic risk scores (PRSs) with the potential to enhance current risk models, but validation is often limited. OBJECTIVE: We set out to validate a set of high-powered BC PRSs for use in a local population using an alternate genotyping platform and to assess PRS algorithm capacity to predict other clinical variables that could improve BC screening and treatments. METHODS: Two published PRS algorithms (313 and 3,820 SNPs) were used to score female subjects in this retrospective case-control study utilizing breast cancer positive cases from the integrated Cancer Repository for Cancer Research and cancer negative controls from the Netherland Twin Register which have been shown to be genetically similar populations. Subject biospecimens were genotyped on the Illumina Global Screening Array followed by imputation, principle component analysis (PCA), and quality control. Phenotypic data was collected using patient-based questionnaires for cases and controls, with additional case electronic medical record data. 409 cases and 1,375 controls and were ultimately scored via ScoreRunner, a custom bioinformatics pipeline. PRS performance was compared using area under receiver operating characteristic curve (AUROC) analysis. Phenotypic data of interest included age at diagnosis, cancer stage at diagnosis, and histopathological variables. Age at diagnosis was evaluated using survival curve analysis with subjects binned by standard deviation. RESULTS: Mean PRS standardized residual scores were significantly different (pth percentile bin starting at approximately age 40 that persisted throughout life. Stage at diagnosis showed a non-significant positive trend with an increasing 3,820 PRS, with no trend seen for histopathological variables with either PRS. CONCLUSIONS: Our study validates the results and predictive performance of the previously published work in a Midwest-American population using an alternative genotyping platform and showed that high PRS predicts earlier BC diagnosis. Assessment of other clinical variables did not reveal significant trends, indicating the specificity and limitations of PRSs for alternate uses and the need for trait specific PRS development.

PrgmNr 3633 - Increased burden of pathogenic variants in cancer predisposition genes in minority Americans with a cancer diagnosis between age 0-26: Use of linked California health registries

View session detail

Author Block: **Q. Feng**[1], E. Nickels[2], I. S. Muskens[1], A. J. de Smith[1], W. J. Gauderman[1], A. Yee[1], S. Feurstein[3], K. McNeely[3], C. N. Ricker[4], T. Mack[1], A. D. Leavitt[5], L. A. Godley[3], J. L. Wiemels[1]; [1]USC Keck Sch. of Med., Los Angeles, CA, [2]Children's Hosp. Los Angeles, Los Angeles, CA, [3]The Univ. of Chicago, Chicago, IL, [4]USC Norris Comprehensive Cancer Ctr., Los Angeles, CA, [5]Univ. of California, San Francisco, San Francisco, CA

Disclosure Block: **Q. Feng:** None.

The causes of most early-onset cancer, including solid and hematologic cancers, are not well defined in minority Americans. While common genetic variants and environmental factors likely contribute to population prevalence of early-onset cancer, there are also rare single-gene syndromes that contribute to familial clustering of disease. The latter can be identified by evaluating family-based cancer concordance and sequencing of predisposition genes. The level of such familial concordance for early-onset cancers has not been assessed previously in non-European ancestral groups.We used the California Cancer and Vital Statistics Registries to evaluate the relative risks of cancer for first-degree relatives of patients diagnosed between ages 0-26. From 1989-2015, we identified 29249 cancer patients (29072 probands, 177 affected family members) and 62863 healthy family members. We calculated the standardized incident ratios (SIRs) of early-onset primary cancers diagnosed in siblings and mothers, compared to the expected age- and gender-specific cancer rates in California. We also performed exome sequencing on 35 sibling pairs with at least one hematologic cancer (leukemia/lymphoma) in one of the pairs. The GATK germline short variant pipeline and PeCanPIE pipeline were used to call and determine the pathogenicity of variants that were shared between the siblings. RFMix was used to analyze the local ancestry of pathogenic variants.In the full sample, there was an increased relative risk of any early-onset cancer for siblings and mothers (SIR=3.32; 95%CI: 2.85-3.85) of an affected probands. Given a proband with solid cancer, both Latinos (SIR=4.98; CI:3.82-6.39) and non-Latino Blacks (NLB) (SIR=7.35; CI:3.36-13.95) exhibited significantly higher relative risk of any cancer in siblings and mothers when compared to non-Latino White subjects (NLW) (SIR=3.02; CI:2.12-4.16). For hematologic cancers, higher familial risk was observed for non-Latino Asian/Pacific Islanders (NLAPI) (SIR=7.56; CI:3.26-14.90) compared to NLW (SIR:2.69; CI:1.62-4.20). Rare pathogenic variants shared between siblings were detected in 5 of the 35 sibling pairs. Among them, 4 were Latino pairs that shared variants in *TP53* and *ATM*; 1 was a NLB pair that shared a variant in *GATA2*. Interestingly, all pathogenic variants reported above (N=5) reside in European ancestry haplotypes despite the higher familial concordance in ethnic minority families.The data support increased risk of familial-associated early-onset cancers among Latino, NLB, NLAPI compared to NLW subjects. Most specific genetic variants that might explain this finding were discovered in Latino families.

PrgmNr 3634 - Inherited *TP53* variants are associated with an increased risk of prostate cancer

View session detail

Author Block: **H. Cheng**[1], J. Powers[2], R. Gulati[3], A. Le[4], E. Ledet[5], E. Van Allen[6], J. Vijai[7], P. Nicolosi[8], R. L. Nussbaum[9], J. E. Garber[10], K. Offit[7], J. Schiffman[11], O. Sartor[5], P. S. Nelson[3], M. F. Walsh[7], C. C. Pritchard[1], K. N. Maxwell[12]; [1]Univ. of Washington, Seattle, WA, [2]Univ. of Pennsylvania, Philadelphia, PA, [3]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [4]Med., Div. of Hematology-Oncology, Univ Pennsylvania, Philadelphia, PA, [5]Tulane Cancer Ctr., Tulane Med. Sch., New Orleans, LA, [6]Boston, MA, [7]Mem. Sloan Kettering Cancer Ctr., New York, NY, [8]Invitae, Palo Alto, CA, [9]San Francisco, CA, [10]Dana Farber Cancer Inst, Boston, MA, [11]Univ of Utah, Salt Lake City, UT, [12]Univ Pennsylvania, Philadelphia, PA

Disclosure Block: **H. Cheng:** None.

Background: Inherited germline *TP53* pathogenic and likely pathogenic mutations (g*TP53*) cause autosomal dominant multi-cancer predisposition including Li-Fraumeni Syndrome (LFS). However, there is no known association of prostate cancer with *gTP53*. In this study, we aimed to determine if gTP53 predisposes to prostate cancer. **Methods:** This multi-institutional retrospective study characterizes prostate cancer incidence in a cohort of LFS males, and g*TP53* prevalence in a prostate cancer cohort. We evaluated the spectrum of g*TP53* variants and clinical features associated with prostate cancer. **Results:** We identified 31 prostate cancer cases among 163 adult LFS males, including 26 of 54 age ≥50y. In 117 LFS males without prostate cancer at the time of genetic testing, six were diagnosed with prostate cancer over a median (IQR) of 3.0y (1.3-8.0) of follow-up, a 25.1-fold increased risk (95% CI 9.2-54.7; PTP53 in 38 of 6,850 males (0.6%) in the prostate cancer cohort, a relative risk 9.1x higher than population controls (95% CI 6.2-13.5; pTP53 prostate tumors had somatic inactivation of the second *TP53* allele. Among g*TP53* prostate cancer cases in this study, the median age at diagnosis was 56y (IQR: 51-62), 42% had Gleason â¥8 tumors, and 30% had advanced disease at diagnosis. **Conclusions:** Complementary analyses of prostate cancer incidence in LFS males, and g*TP53* prevalence in prostate cancer cohorts suggest that g*TP53* predisposes to aggressive prostate cancer. Prostate cancer should be considered as part of LFS screening protocols, and *TP53* considered in germline prostate cancer susceptibility testing.

PrgmNr 3635 - The immuno-genetics of viral antigen response influence glioma susceptibility and survival in a subtype-specific manner

View session detail

Author Block: **G. Guerra**[1], L. Kachuri[1], G. Wendt[1], H. Hansen[1], S. J. Mack[1], A. Molinaro[1], P. Bracci[1], J. Wiencke[1], T. Rice[1], J. E. Eckel-Passow[2], R. B. Jenkins[2], M. Wrensch[1], S. Francis[1]; [1]Univ. of California, San Francisco, San Francisco, CA, [2]Mayo Clinic, Rochester, MN
Disclosure Block: **G. Guerra:** None.
Infections and human leucocyte antigen (HLA) genetic variants have been independently linked to diffuse glioma risk. Recently, several clinical trials have suggested a prognostic benefit to antiviral medications in glioma treatment. In this study we use polygenic risk scores (PRS) to assess the relationship between genetic predictors of antibody response to 7 viral infections and glioma risk and survival.
We constructed PRSs for seroreactivity to each viral antigen using genome-wide significant (p-8) independent variants (linkage disequilibrium r^2 Genetically predicted stronger antigen response to Epstein-Barr virus (EBV) ZEBRA antigen was inversely associated with glioma risk overall (odds ratio (OR)=0.94, p=0.011) and in opposite directions with response to Merkel cell polyomavirus (MCV) L1 antigen in the IDH wild type (OR_{ZEBRA}=0.91, p=0.008 / OR_{MCV}=1.09, p=0.011) and TERT-only (OR_{ZEBRA}=0.89, p=0.011/ OR_{MCV}=1.11, p=0.018) subtypes. Correlation between PRS_{ZEBRA} and PRS_{MCV} (Pearsonâ□□s r =-0.354, p=3.8Ã□10^{-14}) suggests a shared underlying germline mechanism. In line with these findings, HLA-DQA1*03:01, which influences response to EBV ZEBRA (p=1.3Ã□10^{-16}), was associated with glioma overall (OR=1.18, p=3.9Ã□10^{-4}), IDH wild type (OR=1.22, p=0.0014), and TERT-only (OR=1.21, p=0.034) subtypes. PRS risk associations for the EBV EBNA antigen were seen in IDH-mutated (OR=1.09, p=0.04) subtypes, with the largest effect in the IDH-mutated/1p19q co-deleted (OR=1.14, p=0.031) group.
Considering clinical outcomes, we detected associations between PRS_{EBNA} and survival time in patients with IDH mutated/1p19q non-codeleted tumours (HR=0.86, p=0.019 / 615 patients, 222 events), and PRS_{EBNA} and PRS_{ZEBRA} being associated in opposite directions with survival in IDH-mutated/1p19q co-deleted tumours (HR_{EBNA}=0.75, p=0.019 / HR_{ZEBRA}=1.27, p=0.019 / 244 patients, 64 events).
Our study is the first to associate genetically predicted immune response to viruses with glioma risk. We demonstrate that HLA-DQA1*03:01 influences glioma risk in the same glioma subtypes, potentially suggesting a shared genetic architecture. We also provide preliminary evidence of a prognostic value for genetically programmed reactivity to two EBV antigens for patients with IDH-mutated gliomas. Further studies are required to disentangle the complex interactions between the HLA, infections and glioma risk and survival.

PrgmNr 3636 - CDH1 and CTNNA1 variants with incomplete penetrance in Hereditary Diffuse Gastric Cancer Chilean families

View session detail

Author Block: **G. Molina**[1,2], A. Estrada[1], A. Poceros[1], N. Coggins[1], C. Parra[2], O. Torres[3], E. Norero[4,5], A. CorvalÃin[5,6], L. Carvajal-Carmona[1,7]; [1]UC Davis Genome Ctr.. Univ. of California Davis., Davis, CA, [2]LINA. Facultad de Ciencias de la Salud. Univ. de Playa Ancha, Valparaiso, Chile, [3]Servicio de CirugÃa Digestiva. Hosp. Regional Guillermo Grant., ConcepciÃ³n, Chile, [4]Esophagogastric Surgery Unit, Surgery Dept., Hosp. Dr. Sotero del Rio, Santiago, Chile, [5]Advanced Ctr. for Chronic Disease (ACCDiS), Santiago, Chile, [6]Facultad de Med.. Pontificia Univ. CatÃ³lica de Chile,, Santiago, Chile, [7]Dept. of Biochemistry and Molecular Med., Sch. of Med., Univ. of California, Sacramento, USA, Sacramento, CA
Disclosure Block: **G. Molina:** None.
The purpose of this study is to describe two Chilean families presenting Hereditary Diffuse Gastric Cancer (HDGC). Variants of uncertain significance (VUS) were detected in these families. Gastric cancer is the leading cause of cancer-related death in Chile. Worldwide, only ten percent of gastric cancer have a familial aggregation. E-Cadherine 1 gene (CDH1) and CTNNA1 (alpha-1 catenin gene) are the most commonly mutated genes in HDGC. The index patient of the first family had diffuse gastric cancer at 59 years of age. His father and paternal grandfather died of gastric cancer. A germline variant c.88C>A (p.Pro30Thr, rs139866691) was found previously in the index patient. Genotypification of the family members showed that two non-affected sisters presented the same variant, indicating an incomplete penetrance. In the second family, there are eleven members affected with gastric cancer and two with breast cancer, in two following generations. The index patient was diagnosed with a diffuse gastric tumor at 51 years of age. Cancer gene panel sequencing was performed in index patient DNA and variant genotyping in her relatives. A germline variant c.293G>A (p.R98Q, rs746832629) in the CTNNA1 gene was found in the index patient. We found this variant in two non-affected brothers and it was not found in an affected sister, findings that could be explained by incomplete penetrance. Both variants are in conserved domains of the proteins. CDH1 variant codifies for a residue in a conserved loop of the E-Cadherine 1 preprotein domain that is crucial to the normal positioning of the protein at the membrane. Whereas, CTNNA1 variant codifies for a residue in the first alpha-helix of alfa-1-catenin corresponding to the beta-catenin binding domain. These residues are highly conserved in phylogenetic analysis and the variants population in the Genome Aggregation Database are 2 x 10-3 and 7,5x 10-5. Based on this data, we think that both variants classified as VUS, could be pathogenic and causative of the HDGC observed in these families, but showing an incomplete penetrance. We are performing isogenic models and functional analyses of both variants to clarify their pathogenicity. Also, we are doing WES in the samples of two discordant affected sisters from the second family.

PrgmNr 3637 - Saturation Genome Editing of *PALB2* Reveals Functionally Abnormal Missense Variants

View session detail

Author Block: **I. Hill**[1,2], A. Patoski[1,2], K. de Leon[1,2], A. Mehrotra[1,2], N. Smith[1,2], J. Shendure[1,2,3], L. Starita[1,2]; [1]Dept. of Genome Sci., Univ. of Washington, Seattle, WA, [2]Brotman Baty Inst. for Precision Med., Seattle, WA, [3]Howard Hughes Med. Inst., Seattle, WA
Disclosure Block: **I. Hill:** None.
Pathogenic variants in the *PALB2* gene are associated with an increased lifetime risk of developing breast, pancreatic, and ovarian cancer. However, the clinical utility of genetic testing for informing *PALB2*-associated cancer susceptibility is limited by a lack of evidence for interpreting individual variants. In fact, of the 3125 *PALB2* single nucleotide variants (SNVs) in ClinVar, 1850 (60%) are currently annotated as variants of uncertain significance (VUS). The situation is worse for missense variants, of which 1849 of 1874 (99%) are VUS. Multiplexed assays of variant effects can provide valuable evidence for reinterpreting VUS. For example, functional data from Saturation Genome Editing of *BRCA1* can be used as strong evidence that individual variants are pathogenic (PS3) or benign (BS3). Like *BRCA1*, *PALB2* - **P**artner and **L**ocalizer of **B**RCA**2** - which functions in the same double strand break repair pathway as *BRCA1*, is compatible with Saturation Genome Editing.
We performed Saturation Genome Editing on *PALB2* to generate functional data for all possible SNVs across the protein coding region and intron-exon junctions. Thus far, we have data for 2602 variants, spanning the coiled coil domain which is required for BRCA1 binding and the WD40 domain which is required for binding to BRCA2. Functional scores indicate that 176 of 1408 (12.5%) missense variants are indistinguishable from nonsense variants and are likely â□□functionally abnormalâ□□, loss-of-function variants. Our results are consistent with what is known about PALB2 function, with missense variants scoring as functionally abnormal falling in the coiled-coil (10.5%) or WD40 domain (15.6%), and 0% outside of those domains thus far (n.b. there is more missing data outside of the folded functional domains). Our results are also perfectly concordant for 26 variants with published orthogonal functional data. Nonetheless, it should be recognized that the relative lack of classified pathogenic or benign missense variants within *PALB2* presents a unique and challenging problem in validating these data for use in clinical variant interpretation. New strategies may be necessary before these data can be incorporated into variant interpretation frameworks. Finally, recent case-control studies suggest that *PALB2* missense variants do not appear to increase breast cancer risk, however, our data suggests grouping all missense variants for these studies, of which 87.5% are â□□functionally normalâ□□, may have obscured any signal.

PrgmNr 3638 - A scalable deep learning framework for breast cancer prediction using DNA methylation data

View session detail

Author Block: **N. Paul**[1], R. Jansen[1], R. Gomes[2], N. He[2], A. Huber[2]; [1]North Dakota State Univ., Fargo, ND, [2]Univ. of Wisconsin, Eau Claire, WI

Disclosure Block: **N. Paul:** None.

DNA methylation is a key epigenetic modification that can modulate gene expression to influence the cell functionality. This process affects tumor suppressor genes and oncogenes to lead to cancer. DNA methylation can be measured by high-throughput sequencing technology. However, these technologies cannot determine complete CpG coverage and hence prediction of missing methylation states are critical for complete genome-wide analysis. The main challenge for prediction is the high dimensionality and complexity of the methylation data. With high throughput bulk sequencing techniques being able to read over 850,000 methylation markers (CpGs) across the human genome, a need for a more hybrid and scalable approach to interpret these datasets is essential. To address this dimensionality, we propose a two-step method of feature selection and cancer prediction using state-of-the-art machine learning algorithms. The feature selection stage utilizes several classification and clustering algorithms that can select a subset of methylation markers contributing towards gene expression. In this work, we apply our method using breast cancer methylation data from the publicly available Cancer Genome Atlas Program (TCGA). In the first stage, Principal Component Analysis, Random Forest, and ANOVA F-test are applied to select the top CpG markers from an Illumina 450K array that could be the most important cancer related features. The degree of similarity between selected CpG markers is also analyzed using Correlated Feature Elimination and some of the highly correlated features would be documented and removed prior to classification. A Convolutional Neural Network (CNN) is implemented to develop a diagnostic prediction model for cancer detection from the selected CpG markers. Preliminary work on a subset of 20 samples shows promising results. The proposed CNN architecture was able to predict cancer samples with an accuracy of 100% using a subset of 77 most important CpG markers from a total of 487,177 markers. To validate the efficacy of the proposed approach, the algorithm is applied on 1188 samples. Our proposed approach consisting of Feature Selection stage would make this model scalable and faster while Deep Learning would improve accuracy of the prediction model and allow capture of non-linear relationships. Results and weights from the trained deep learning model will be shared with the cancer research community to facilitate transfer learning. This will ensure that the cancer research community gets access to a deep learning framework that already has preliminary knowledge from past experiments.

PrgmNr 3639 - Cell-type specific methylation changes from bulk whole genome bisulfite sequencing of matched primary and recurrent high grade serous ovarian cancers indicates stability of methylation landscape through tumor progression and chemoresistance

View session detail

Author Block: **N. Gull**[1], P-C. Peng[2], S. G. Coetzee[3], T. Silva[4], J. Plummer[3], B. Davis[1], S. Chen[5], K. Lawrenson[6], B. Berman[7], B. Karlan[8], S. A. Gayther[3], M. R. Jones[9]; [1]Ctr. for Bioinformatics and Functional Genomics, Dept. of BioMed. Sci., Cedars-Sinai Med. Ctr., Los Angeles, CA, [2]Cedars-Sinai Med. Ctr., Sunnyvale, CA, [3]Cedars Sinai Med. Ctr., Los Angeles, CA, [4]Div. of Biostatistics, Dept. of Publ. Hlth.Sci., Univ. of Miami, Miller Sch. of Med., Miami, FL, [5]Culver City, CA, [6]Cedars-Sinai Med. Ctr., Los Angeles, CA, [7]Dept. of Dev.al Biology and Cancer Res., Inst. for Med. Res. Israel-Canada, Hebrew Univ.-Hadassah Med. Sch., Jerusalem, Israel, [8]Dept. of Obstetrics and Gynecology, Jonsson Comprehensive Cancer Ctr., David Geffen Sch. of Med., UCLA, Los Angeles, CA, [9]Cedars Sinai Med Ctr, Los Angeles, CA
Disclosure Block: **N. Gull:** None.

The role of DNA methylation changes in high grade serous ovarian carcinoma (HGSOC) is largely unknown. We performed whole genome bisulfite sequencing (WGBS) and RNA-seq in paired primary and recurrent tumors 28 HGSOC patients (62 total). Genome wide methylation and transcriptomic features were largely preserved between primary and recurrent tumors from the same patient ($P=7.16x10^{-7}$ in *BRCA*,$1.41x10^{-3}$ in non-*BRCA*). We did not identify any differentially methylated regions (DMRs) or consistent loss/gain of methylation at partially methylated domains (PMDs) between primary and recurrent tumors across the cohort. Within-patient paired analysis of DMRs identified an average of 659 DMRs between the primary and recurrent tumors from *BRCA* carriers, which was significantly higher (P=0.004) than the 388 average DMRs identified from the paired analysis of primary and recurrent tumors from non-carriers. Tumors from *BRCA* carriers displayed high levels of heterogeneity, likely driven by homologous recombination repair deficiency and a high burden of structural variation, and may be related to poorer survival observed in *BRCA* non-carriers (P=0.0056). Comparing tumors from *BRCA* carriers with non-carriers we identified 135 significant DMRs, with a trend for hypermethylation in DMRs and PMDs in tumors from *BRCA* carriers (P=0.001). These DMRs are depleted within CpG islands near transcription start sites and active regulatory regions in related cell types. These changes were also observed in RNA-seq analysis, where we identified differentially expressed genes between *BRCA* carriers and non-carriers that include the known driver amplification CCNE1 and genes enriched in immune pathways and pathways that maintain stemness and cell differentiation. We applied an in silico single cell WGBS approach using the tool CluBCpG to identify cell types and their frequency in each tumor from bulk WGBS. Results from read grouping analysis support the stability of the methylome between primary and recurrent tumors, even at a cell-type specific level. As PMDs are cell type specific, we focused further analysis on these regions. A larger number of cell-type specific populations were estimated from read grouping in PMDs than non-PMDs ($P=1.82x10^{-9}$) supporting their cell type specific patterns of methylation. Ongoing analysis will compare specific cell type populations and their frequencies between tumors. The conservation of methylation signatures in progressive, chemoresistant HGSOC tumors has not been previously reported, and suggests that chemoresistance may be established early in tumor progression, potentially in small cell populations and via hypomethylation at PMDs.

PrgmNr 3640 - Exome-wide rare variant association study in lymphoid cancers

View session detail

Author Block: **S. Ralli**[1], S. Jones[2], S. Leach[1], A. R. Brooks-Wilson[3]; [1]Genome Sci. Ctr., Vancouver, BC, Canada, [2]BC Cancer, Vancouver, BC, Canada, [3]BritishColumbia Cancer Res Ctr, Vancouver, BC, Canada

Disclosure Block: **S. Ralli:** None.

Genome-wide association studies have revealed common genetic variants associated with diverse lymphoid cancers. However, these variants explain only a portion of the heritability of these cancers. Rare variants are likely to play a role in disease causation. We aim to identify rare germline variants associated with lymphoid cancers using exome sequencing. 39 multiple-case lymphoid cancer families with 86 cases were exome sequenced, jointly called, and filtered using variant quality score recalibration. Cases with the earliest onset or rarest type of lymphoid cancer were selected for rare variant association analysis from each family. Ethnicities for these cases were determined by PCA using KING; 38 cases had European ethnicity. Read depths filter â¥ 10 in 90% of samples were applied to the 38 cases and the genomes of 7718 controls from GnomAD of European ethnicity. Variants of cases were filtered to retain those with (1) CADD â¥ 20, (2) variant consequence of splice acceptor, splice donor, stop gained, frameshift, stop lost, and start lost variant, (3) allele frequency â¤ 0.001 or novel variant with no allele frequency in public databases and (4) genotype 0/1 or 1/1 in at least one case. Variants of controls were filtered using criteria (2) and (3). Burden analysis for rare variant association was performed using TRAPD under a dominant model using single nucleotide variants and indels. TRAPD performs a one-sided Fisher's exact test to analyze if there is a significantly higher burden of variants in cases than controls for each gene. An inflation factor of 1.9 was observed. 6811 genes were analyzed. 189 genes had nominal p-values â¤ 0.05; 159 withstood adjustment for the inflation factor. Of these, *FANCD2* (p-value 0.009) and *SYK* (p-value 0.009) are implicated in causing leukemia and peripheral T-cell lymphoma in the Cosmic cancer gene census. Case with *FANCD2* splice donor variant had another affected individual in the family who was sequenced and they both share the variant. After multiple testing corrections, however, these genes did not remain significant.

PrgmNr 3641 - Identifying non-coding drivers of ovarian cancer by converging germline variants and somatic mutations

View session detail

Author Block: **P-C. Peng**[1], J. Tyrer[2], B. Davis[1], S. Chen[1], F. S. Dezem[1], S. Kar[3], J. Plummer[1], Ovarian Cancer Association Consortium, A. Gusev[4], S. Knott[1], M. L. Freedman[4], P. Pharaoh[2], K. Lawrenson[1], S. A. Gayther[1], M. R. Jones[1]; [1]Cedars-Sinai Med. Ctr., Los Angeles, CA, [2]Univ. of Cambridge, Cambridge, United Kingdom, [3]Univ. of Bristol, Bristol, United Kingdom, [4]Dana-Farber Cancer Inst., Boston, MA
Disclosure Block: **P. Peng:** None.
There are numerous examples where germline and somatic variants target the same genes in specific cancers, indicating a common biology drives the development of both heritable and sporadic cancers. So far, these studies have focused on protein-coding genetic variants. Advances in whole genome sequencing (WGS) methods, have enable the identification of tens of thousands of non-coding germline and somatic variants of unknown function and significance. We hypothesize that a proportion of non-coding germline risk variants and somatic mutations co-locate with regulatory elements (REs) that regulate the expression of specific genes and/or gene regulatory networks. We have integrated germline risk variants identified from genome-wide association studies (GWAS) of ovarian cancer (up to 26,151 cases, 105,724 controls) and somatic mutation data from WGS analysis of ovarian tumors (n > 200) with disease specific epigenomic and transcriptomic data from ENCODE, Roadmap and in-house profiling studies, to identify REs where germline and somatic variants co-localize, which indicates they may be drivers of disease pathogenesis.
We observed significant enrichment of germline variants in transcription factor binding sites (TFBS) in ENCODE ($p = 0.03$) and ovarian cancer-specific active promoters ($p = 0.009$). Somatic mutations mostly occurred in super enhancers (205/8642 mutated, 2.37%) and ovarian cancer-specific active regions (26/1099 mutated, 2.37%). We also found 60 significantly mutated disease associated active REs that harbor putative causal risk variants for ovarian cancer. We predicted the target genes of these REs based on: (1) distance to promoters; (2) association between H3K27ac activity and target gene's expression; (3) long-range chromatin interactions from ovarian cancer cell lines. We identified co-localized variants within; an active enhancer at *2q14* associated with *PAX8*, a lineage-specific TF thought to regulate cell migration and metastasis in ovarian cancer; the promoter of the *TERT* gene at *5p15.33*; and an active enhancer at *1p34.3* linked to *RSPO1*. Pathway analysis mapped putative non-coding drivers to pathways involved in *TERT* activation in cancer ($p = 0.01$), *PAX8* targets in thyroid dysgenesis ($p=0.03$), and regulation of response to DNA damage stimulus ($p=0.03$). In conclusion, this study suggests that germline and somatic non-coding genetic alterations co-localize in REs to indicate novel pathways and biological drivers of ovarian cancer pathogenesis

PrgmNr 3643 - Spatial molecular profiling of Triple-Positive Ductal Carcinoma In Situ of the Breast using Visium Spatial Gene Expression for FFPE

View session detail

Author Block: **S. R. Williams**[1], M. Turkekul[1], A. Jurek[1], M. Mignardi[1], P. Mielinis[1], X. Huang[2], M. Stoeckius[1], V. Giangarra[1], N. I. Weisenfeld[1], J. Chell[1]; [1]10x Genomics, Pleasanton, CA, [2]Univ. of Washington, Seattle, WA

Disclosure Block: **S.R. Williams:** None.

Triple-positive ductal carcinoma in situ of the breast (TPDCIS) is a pre-neoplastic lesion that can give rise to invasive ductal breast carcinoma. Such lesions routinely exhibit elevated levels of HER2, ER, and PR but the underlying phenotype and survival rates vary. To investigate the tumorigenic potential of TPDCIS, we examined the tissue microenvironment of two TPDCIS samples, characterizing the tissue distribution of cellular gene expression using spatial transcriptomics

Spatial transcriptomics technology combines histological techniques with the massive throughput and discovery power of RNA sequencing, adding a powerful toolset to pathological examination. However, standard clinical workflows collect formalin-fixed paraffin-embedded (FFPE) tissue, which can significantly damage molecules such as RNA, making investigation at the transcriptomic level intractable.

We assessed the profiles of two TPDCIS samples using the10x Genomics Visium Spatial Gene Expression Solution for FFPE tissue with a new RNA Templated Ligation (RTL) approach that probes native RNA. Tissue sections on a Visium gene expression slide were H&E stained, imaged, and spatially barcoded to generate libraries. Sequencing data from the libraries were visualized for spatial and quantitative gene expression profiles.

Spatial profiles aligned with pathologist annotation demarcating the tumor, stromal, and immune compartments as well as with staining for HER2, PR, and ER. These data were paired with single nucleus RNA-seq (Chromium Single Cell 3â□□ assay), generating cell type expression profiles for estimating spot level cell type proportions and gaining insight into cell-type co-localization. New computational methods were used to infer copy number variation (CNV) at the spot level, elucidating the spatial relationship between CNVs and intra-tissue phenotypes.

These results demonstrate that gene expression profiling of FFPE tissues using the Visium platform complements traditional histopathological methods. TPDCIS gene expression profiles at both the single nucleus and spatial level provide comprehensive tissue architecture insights. This knowledge enables unprecedented understanding of tumor biology, disease progression, potential predictive biomarkers, and development of therapeutic targets.

PrgmNr 3644 - Tracing intratumoral genetic heterogeneity using genetic barcoding in small cell lung cancer

View session detail

Author Block: **H. Wollenzien**[1,2], Y. A. Tecleab[2], M. Mukherjee[2], M. S. Kareta[2]; [1]Univ. of South Dakota, Sioux Falls, SD, [2]Sanford Res., Sioux Falls, SD
Disclosure Block: **H. Wollenzien:** None.
Small Cell Lung Cancer (SCLC) is often a heterogeneous tumor, where multiple populations of phenotypically different cells exist and contribute differentially to tumor dynamics. This tumor is characterized by a very low 5- year survival rate, high metastatic rates, and rapid acquisition of chemoresistance. Beyond first-line chemotherapy, targeted treatment options are scarce, and none take in to consideration the complex genetics of the tumor. The heterogeneous nature of this tumor makes it difficult to study and to treat, as current attempts to understand tumor dynamics use a bulk approach that averages the contribution of all cells within a tumor. This work uses a novel cellular barcoding lineage tracing approach combined with single-cell RNA sequencing (scRNA-seq) to understand tumor heterogeneity and evolution of SCLC in both a xenograft model of human SCLC cell lines and an in vivo system using a genetically modified mouse model with alterations in the same driver genes as the majority of humans with SCLC. Tumors are sampled pre-growth, post growth and metastasis, and pre- and post- chemotherapy. By paring the barcode populations that exist in samples over time, in metastasis, or in chemoresistance, a clonal evolution map of growth, metastasis, and chemoresistance can be generated, based on shared barcodes and convergent or divergent transcriptomic profiles. In validation of scRNA-seq data, an emphasis will be placed on genes that may already be targetable by approved therapeutics. This work will identify the genes responsible for tumor dynamics, develop the in vivo cellular barcoding approach, and lead to the identification and investigation of therapeutic targets for human SCLC.

PrgmNr 3645 - Transcriptome-wide association study identifies novel genes associated with the immune traits in cancer

View session detail

Author Block: **P. Middha Kapoor**[1], R. W. Sayaman[2], J. L. Nierenberg[3,1], M. Saad[4,5], V. Thorsson[6], B. Davide[7,8,9], E. Ziv[10]; [1]Dept. of Med., Univ. of California San Francisco, San Francisco, CA, [2]Dept. of Lab. Med., Helen Diller Family Comprehensive Cancer Ctr., Univ. of California San Francisco, San Francisco, CA, [3]Univ. of California San Francisco, San Francisco, CA, [4]Qatar Computing Res. Inst., Hamad Bin Khalifa Univ., Doha, Qatar, [5]NeuroSci. Res. Ctr., Faculty of Med. Sci., Lebanese Univ., Beirut, Lebanon, [6]Inst. for Systems Biology, Seattle, WA, [7]Coll. of Hlth.and Life Sci., Hamad Bin Khalifa Univ., Doha, Qatar, [8]Res. Branch, Sidra Med., Doha, Qatar, [9]Dept. of Internal Med. and Med. Specialties, Univ. of Genoa, Genoa, Italy, [10]Inst. for Human Genetics, Helen Diller Family Comprehensive Cancer Ctr., Univ. of California San Francisco, San Francisco, CA

Disclosure Block: **P. Middha Kapoor:** None.

Background: Immune infiltration in solid tumors is a strong predictor of improved survival in many tumor types. We have recently demonstrated the heritability of immune infiltration components in solid tumors by using germline genetic data from The Cancer Genome Atlas (TCGA). We also identified over 20 individual loci that are associated with over 10 immune signatures including genes that are involved in autoimmune disorders and genes that are proposed targets of immunotherapy. Here, we sought to identify additional genes associated with variation in the immune microenvironment. We performed transcriptome-wide association study (TWAS) to predict immune signatures in the pan-cancer analyses of 30 non-hematological cancers in TCGA. Methods: We integrated the results from the pan-cancer genome-wide association study with large-scale expression quantitative trait loci (eQTLs) from whole blood, spleen, and EBV-transformed lymphocytes tissues in GTEx (version 8). We conducted a TWAS using the Summary-MulTiXcan approach. We used an false discovery rate of ZBTB8OS, RAB43, TNNT1, TSSK3, ZNF134, SAP30B, LRFN3, UPK3BL, EPHB6, GKAP1, NT5C, and *RP11-290F24.6*) whose genetically predicted expression was associated with different immune signatures in tumors. One of these genes, *EPHB6* was found to be inversely associated with a Th1 enrichment ($p=2.2x10^{-6}$). EphB6 plays a crucial role in T-cell activation with EphB6-deficient mice displaying reduced activation, phosphorylation and recruitment of the T cell signaling molecules. *RAB43* was found to be inversely associated with activated a signature of natural killer cells ($p=3.8x10^{-7}$). RAB proteins are involved in signal transduction pathways regulating cell invasion, cell apoptosis and innate immune response, particularly in gliomas, leading to poor clinical outcomes. Conclusion: This is the first TWAS investigating the relationship between genetically predicted gene expression and immune traits. Of these, *EPHB6,* and *RAB3* are strong candidates for a mechanistic role in modifying the immune response to tumors.

PrgmNr 3646 - Epigenome-wide association study (EWAS) for soluble CD14 level in African Americans from the Jackson Heart Study

View session detail

Author Block: **K. Ferrier**[1], N. E. Olson[2], A. Reiner[3], E. M. Lange[4], L. A. Lange[5]; [1]Univ. of Colorado Denver, Aurora, CO, [2]PerkinElmer, Seattle, WA, [3]Univ of Washington, Seattle, WA, [4]Univ North Carolina, Chapel Hill, NC, [5]Univ Colorado Denver, Aurora, CO
Disclosure Block: **K. Ferrier:** None.
The human monocyte differentiation antigen CD14 is a pro-inflammatory cytokine that has been implicated in the pathogenesis of a number of disease processes including inflammation, metabolic disease, tumor development, and atherosclerosis. In a separate analysis, we showed soluble CD14 (sCD14) is associated with incident coronary heart disease, heart failure, and all-cause mortality independently of established cardiovascular disease risk factors in African Americans (AAs) from the Jackson Heart Study (JHS). We also identified novel associated genetic variants near the *CD14* locus. To understand the epigenetic influences on CD14, we performed an EWAS for sCD14 levels in n=1613 JHS participants with available Illumina EPIC array data (n=721,757 methylation sites) measured from whole blood. Methylation values were normalized and adjusted for batch effects. Epigenome-wide association analysis was performed using a linear regression model. Analyses were adjusted for age, sex, estimated cell counts (Houseman method), smoking status and 10 genetic ancestry principal components. We observed four methylation sites that reached epigenome-wide significance (p-8) for association with sCD14 levels after multiple-testing correction: cg27161585 near *TMC06* (near the *CD14* gene), cg26470501 near *BCL3*, and cg03396765 and cg15627993 near *LDLRAD4*. Conditional analyses of cg27161585 by a previously identified sCD14-associated SNP in the *CD14* gene region (rs75652866) completely attenuated the association between the methylation site and sCD14 level, suggesting that rs75652866, or a variant in high linkage disequilibrium with it, may regulate CD14 expression. In conclusion, this study identified several novel possible regulatory regions associated with sCD14 levels in AAs previously not identified in GWAS. Methylation levels near *CD14* do not appear to provide additional predictive value for sCD14 levels beyond what is explained by identified associated genetic variants. Normal 0 false false false EN-US X-NONE X-NONE /* Style Definitions */ table.MsoNormalTable {mso-style-name:'Table Normal'; mso-tstyle-rowband-size:0; mso-tstyle-colband-size:0; mso-style-noshow:yes; mso-style-priority:99; mso-style-parent:''; mso-padding-alt:0in 5.4pt 0in 5.4pt; mso-para-margin:0in; mso-pagination:widow-orphan; font-size:12.0pt; font-family:'Calibri',sans-serif; mso-ascii-font-family:Calibri; mso-ascii-theme-font:minor-latin; mso-hansi-font-family:Calibri; mso-hansi-theme-font:minor-latin; mso-bidi-font-family:'Times New Roman'; mso-bidi-theme-font:minor-bidi;}

PrgmNr 3647 - Exploring the genetic architecture of spontaneous coronary artery dissection using whole genome sequencing

View session detail

Author Block: **I. Tarr**[1], S. Hesselson[1], S. Iismaa[1,2], E. Rath[1,2], S. Monger[1], M. Troup[1], K. Mishra[1], C. M. Y. Wong[1], P-C. Hsu[1], K. Junday[1], D. Humphreys[1], D. Adlam[3], A. A. Baranowska-Clarke[3], S. E. Hamby[3], K. Carss[4], T. R. Webb[3], N. J. Samani[3], M. Bax[1], L. McGrath-Cadell[1,2], J. C. Kovacic[1,2,5,6], S. L. Dunwoodie[1,2], D. Fatkin[1,2,6], D. W. M. Muller[1,2,6], B. M. Graham[1,2,6], E. Giannoulatou[1,2]; [1]Victor Chang Cardiac Res. Inst., Sydney, Australia, [2]Univ. of New South Wales, Sydney, Australia, [3]Univ. of Leicester, Leicester, United Kingdom, [4]Ctr. for Genomics Res., AstraZeneca, Cambridge, United Kingdom, [5]Icahn Sch. of Med. at Mount Sinai, New York, NY, [6]St Vincentâ□□s Hosp., Sydney, Australia
Disclosure Block: **I. Tarr:** None.

Spontaneous coronary artery dissection (SCAD) is a cause of acute coronary syndrome that predominantly affects women. Its pathophysiology remains unclear but connective tissue disorders (CTDs), vasculopathies, and other conditions such as Fragile-X have been observed in SCAD patients. That SCAD has a genetic component is accepted, yet few genes have been robustly implicated. We sought to clarify the genetic etiology of SCAD using targeted and genome-wide methods in sporadic cases to identify both common and rare disease-associated variants.

A cohort of 91 unrelated sporadic SCAD cases were investigated for rare, likely pathogenic variants in genes associated with either SCAD, CTDs, or vasculopathies. New candidate genes were sought using rare variant collapsing analysis relative to 1127 healthy, elderly controls and through identification of novel loss-of-function variants in genes intolerant to such variation. Pathway analysis was used to investigate common themes across potential novel candidate genes. Candidate genes were validated in an independent cohort of 384 SCAD cases. The Fragile X-associated short tandem repeat (STR) was assessed. Finally, two SCAD polygenic risk scores (PRSs) were used to assess the contribution of common variants.

We identified 10 cases with at least one rare, likely disease-causing variant in CTD-associated genes, although only one had a significant CTD phenotype. No genes were significantly associated with SCAD in our genome-wide rare variant collapsing analysis, however, pathway analysis of 24 genes harboring novel loss-of-function variants implicated the TGF-Î² signaling pathway (adjusted p-value = 0.0005). Validation of these 24 genes in 384 independent SCAD cases identified one further case who carried a different novel loss-of-function variant in the gene ACACA. No expansions of the Fragile-X STR were found. Sporadic SCAD cases showed elevated SCAD risk compared to controls for both PRSs (p-value range: 1.372e-09 - 2.052e-10).

Our analysis included screening for single nucleotide variants, insertion/deletions, structural variants of all sizes, splice-altering variants including deep intronic variants, and STR expansions, making this the first study to systematically interrogate WGS data for all types of rare variation to identify potential causes of SCAD. We showed that SCAD shares some genetic overlap with CTDs, both in a targeted analysis of known CTD genes and in potential candidates identified in this study - all despite the absence of any major CTD phenotype in most of the cohort. Consistent with a complex genetic architecture, SCAD patients also have a higher burden of common SCAD variants than controls.

PrgmNr 3648 - Gene expression analysis of lipid traits in multi-ethnic population

View session detail

Author Block: **Z. Du**[1], R. Joehanes[2], Y. Wang[3], T. Huan[4], L. A. Cupples[3], A. Reiner[5], K. E. North[6], C. S. Carlson[1], N. Heard-Costa[7], U. Peters[1], D. Levy[8], G. M. Peloso[9], C. Kooperberg[1]; [1]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [2]NIH, Rockville, MD, [3]Boston Univ. Sch. of Publ. Hlth., Boston, MA, [4]NHLBI-Framingham heart study, Framingham, MA, [5]Univ. of Washington, Seattle, WA, [6]Univ. of North Carolina, Chapel Hill, NC, [7]Boston Univ. Sch. of Med., Boston, MA, [8]NHLBI/NIH, Framingham, MA, [9]Boston Univ., Boston, MA

Disclosure Block: **Z. Du:** None.

Backgrounds Blood concentration of lipids, including total cholesterol (TC), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides (TG) are well-established risk factors for cardiovascular disease (CVD). Gene expression can offer key information for understanding the regulatory and functional mechanisms underlying phenotypic differences among individuals. In the present study, we examined associations between blood RNA-seq and lipid traits to identify genes and pathways involved in regulating blood lipid levels. **Methods** A two-stage association study was performed, with a discovery set in Women's Health Initiative (WHI) and Framingham Heart Study (FHS), including a total of 2383 European-ancestry subjects and 480 non-European subjects, and validation in a second set of 1,189 FHS participants of European-ancestry. The association between transformed RNA-seq counts of each gene with lipid traits was examined using linear regression adjusting for age, genetic PCs, surrogate variables, complete blood counts, lipid-modifying medicines, and other cohort-specific covariates (i.e, sex in FHS). Fixed-effect meta-analysis was used to combine results. Bonferroni adjustment was used to correct multiple tests. Gene and pathway enrichment analyses were performed for replicated differentially expressed genes (DEGs) using R package clusterProfiler. Sensitivity analyses were performed by adding more potential confounders, including BMI, smoking and alcohol drinking, into the model. Causal effects of the detected DEGs will be estimated using Mendelian randomization.
Results A total of 334, 14, 61 and 370 genes were identified in the discovery and replicated to be differentially expressed for blood HDL-C, LDL-C, TC and TG levels, respectively; among which 260, 10, 46 and 312 were not in previously identified GWAS risk *loci* (>250 kb). Enrichment analyses showed overrepresentation in biologic processes including regulation of lipid and cholesterol metabolic process for HDL and TC, Fc receptor signaling pathway for HDL, heart process pathway for LDL and HDL, neutrophil activation for TG, and reproductive system for TC.
Conclusion Our results identify novel genes and pathways related to blood lipids levels, providing clues for future research of biological mechanisms of population lipid diversity and potential therapeutic targets.

PrgmNr 3649 - Investigating a potential causal relationship between maternal blood pressure during pregnancy and future offspring cardiometabolic health

View session detail

Author Block: **G. Wang**[1], L. Bhatta[2], G-H. Moen[3,2,1,4], L-D. Hwang[1,5], J. P. Kemp[1,6,5], T. A. Bond[6,1,4], B. Ã svold[2,7], B. Brumpton[2,8,9], D. M. Evans[1,5,6], N. M. Warrington[1,2,5]; [1]The Univ. of Queensland Diamantina Inst., The Univ. of Queensland, Brisbane, Australia, [2]K.G. Jebsen Ctr. for Genetic Epidemiology, Dept. of Publ. Hlth.and Nursing, NTNU, Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [3]Inst. of Clinical Med., Faculty of Med., Univ. of Oslo, Oslo, Norway, [4]Population Hlth.Sci., Bristol Med. Sch., Univ. of Bristol, Bristol, United Kingdom, [5]Inst. of Molecular BioSci.s, The Univ. of Queensland, Brisbane, Australia, [6]Med. Res. Council Integrative Epidemiology Unit, Univ. of Bristol, Bristol, United Kingdom, [7]Dept. of Endocrinology, Clinic of Med., St. Olavs Hosp., Trondheim Univ. Hosp., Trondheim, Norway, [8]Clinic of Med., St. Olavs Hosp., Trondheim Univ. Hosp., Trondheim, Norway, [9]HUNT Res. Ctr., Dept. of Publ. Hlth.and Nursing, NTNU, Norwegian Univ. of Sci. and Technology, Levanger, Norway
Disclosure Block: **G. Wang:** None.

Background: Observational epidemiological studies have reported that higher maternal blood pressure during pregnancy is associated with increased future risk of offspring cardiometabolic disease. However, it is unclear whether this association represents a causal relationship through intrauterine mechanisms.

Methods: We used a Mendelian randomization (MR) framework to examine the potential causal relationship between unweighted maternal genetic scores for systolic blood pressure (SBP) and diastolic blood pressure (DBP), and a range of cardiometabolic risk factors in the offspring of up to 29,708 genotyped mother-offspring pairs from the UK Biobank (UKB) and the TrÃ¸ndelag Health (HUNT) studies. The cardiometabolic risk factors included in our analysis were SBP, DBP, body mass index, lipid profile (i.e. apolipoprotein A, apolipoprotein B, total cholesterol, low-density lipoprotein cholesterol, lipoprotein A, high-density lipoprotein cholesterol, and triglycerides), glycaemic biomarkers (i.e. non-fasting glucose, glycated haemoglobin, and insulin-like growth factor 1) and other relevant traits (i.e. C-reactive protein and urate). We conducted similar analyses in up to 21,423 father-offspring pairs from the same cohorts as a negative control for postnatal effects. We also calculated the statistical power to detect maternal genetic effects on offspring cardiometabolic risk factors conditional on offspring genotype.

Results: We did not detect any association between maternal (or paternal) unweighted genetic scores for blood pressure and offspring cardiometabolic outcomes in the meta-analysis of UKB and HUNT analyses after adjusting for offspring genotypes at the same loci. Power calculations indicated that we had â¥80% power to detect a maternal genetic effect that explained as little as 0.035% of the variance in the offspring cardiometabolic traits with 29,708 mother-offspring pairs (two-tailed Î± = 0.05).

Conclusions: We find little evidence to support the notion that normal variation in maternal blood pressure is a major causal risk factor for adverse offspring cardiometabolic outcomes in later life. These results complement the findings from conventional epidemiological studies, and indicate that offspring genetic effects or/and other environmental factors predominantly increase offspring cardiometabolic disease risk.

PrgmNr 3650 - Investigating exposure mediated genetic scoring for coronary artery disease

View session detail

Author Block: **D. Adams**[1], W. Reay[2], M. Cairns[3]; [1]Univ. of Newcastle, Callaghan, Australia, [2]Univ. of Newcastle/Hunter Med. Res. Inst., Callaghan, Australia, [3]The Univ. of Newcastle, Callaghan, Australia
Disclosure Block: **D. Adams:** None.
Coronary artery disease (CAD) is significantly heritable and remains one of the leading causes of death worldwide. Current risk prediction methods such as polygenic risk scoring (PRS) may be useful to stratify individuals at risk of CAD but do not indicate actionable interventions that could reduce disease severity. We developed two novel methods; the instrumental variable risk score (IVS) and individual Mendelian randomisation scores (IMR) to investigate the genetic heterogeneity in the manifestation of known CAD risk factors (lipids - triglycerides, low density lipoprotein and total cholesterol). The IVS sums the effect of individual SNPs as genetic instrumental variables that represent the predicted effect of exposures (lipids) on CAD using summary statistics from genome wide association studies. Thus, representing the additive effect of the genetically predicted exposure on CAD within the individual. The IMR method has a similar aim but utilises the individualâ□□s genotypes for IV SNPs as input into several Mendelian randomisation models with different underlying assumptions regarding IV validity. We investigated the relationship between these scores and CAD in the UK Biobank (UKBB) and gene expression in the GEUVADIS cohort. We found IVS and IMR that were significantly associated with the odds of CAD and remained significant for a subset of scores upon covariation for CAD PRS and a polygenic score for the lipid exposure. Interestingly, we found that approximately 22% of the UKBB cohort had a non-significant IMR estimate for the effect of triglycerides on CAD, suggesting there is heterogeneity in this relationship. The addition of IVS and IMR scores for causal exposures can improve the prediction of coronary artery disease risk from genomic information and measured traits, however, more investigation is needed to refine and investigate the broader utility of these approaches.

PrgmNr 3651 - Investigation of eQTLs in the *CELSR2/ PSRC1/SORT1* gene cluster for independent effects in the liver and skeletal muscle that influence LDL levels and energy expenditure in American Indians

View session detail

Author Block: **K. Bandesh**, P. Piaggi, S. Kobes, R. Hanson, C. Bogardus, L. Baier; Phoenix Epidemiology and Clinical Res. Branch, Natl. Inst. of Diabetes and Digestive and Kidney Diseases, NIH, Phoenix, AZ

Disclosure Block: **K. Bandesh:** None.

American Indians (AI) living in southwestern USA have unique genetic architecture and show a heritable predisposition to obesity and type 2 diabetes. To identify genetic variants that contribute to these diseases, we conducted genome-wide association studies for various traits using genotypes derived from a custom Axiom array specific for AI (captures 92% of all variation with minor allele frequency â¥ 0.05). Analysis of fasting LDL levels measured in 5,205 AI identified rs12740374 (G/T) as robustly associated with LDL (P= 1 Ã 10Â$^{-22}$; effect (Î²)= 8% lower per copy of minor allele, T) and total cholesterol (P= 9 Ã 10Â$^{-15}$; Î²= 4% lower). An expression-QTL (eQTL) analysis in the GTEx database showed that rs12740374 associates with the expression of *CELSR2* (P= 10Â$^{-34}$) and neighboring genes *PSRC1* (P= 10Â$^{-37}$) and *SORT1* (P= 10Â$^{-54}$) in liver (N= 208). It is reported that rs12740374 alters the hepatic expression of the SORT1 gene by creating a C/EBP transcription factor binding site. However, the database also showed that rs12740374 is a very strong eQTL for *CELSR2* (P= 10Â$^{-76}$) and *PSRC1* (P= 10^{-12}) in skeletal muscle (N= 706). Therefore, we looked for additional phenotypes associated with this variant which may have a physiologic role in muscle and found that rs12740374 associates with resting metabolic rate (RMR) measured by a ventilated hood system (N= 509; P= 0.003; Î²= -0.13). A major determinant of human metabolic rate is muscle mass. We speculated rs12740374, which has a known role for LDL in liver, maybe tagging a functional variant in the muscle that underlies association with energy expenditure, and observed that rs6670347 (T/C) (r^2= 0.98) located in *CELSR2* intron is similarly associated with LDL and RMR. We did eQTL analysis in AI skeletal muscle (N= 202) and adipose tissue samples (N= 192) and noticed that rs6670347 contributes to higher *CELSR2* expression in muscle (P= 2 Ã 10Â$^{-7}$, Î²= 0.61 SD units per copy of minor allele, C). This eQTL appears to be muscle-specific as no association was detected in studied adipose tissue or other non-muscle tissues reported in the GTEx portal. Further *in-silico* analysis revealed that rs6670347 forms the core binding motif of the NR3C1 transcription factor, and the C allele enables a stronger binding. NR3C1 encodes a receptor for glucocorticoids, which are key regulators of muscle mass, and their prolonged exposure induces atrophy. Glucocorticoid resistance is shown to decrease *CELSR2* expression in human leukemia cells and its inactivation in mice reduces muscle atrophy, which could explain its role in altering RMR. Our findings indicate that different variants in *CELSR2/ PSRC1/SORT1* gene cluster may have tissue-specific effects on diverse metabolic traits.

PrgmNr 3652 - Methylome-wide association analysis identified novel CpG sites for lipid concentrations in ancestral-diverse populations from the PAGE Study

View session detail

Author Block: **Y. Hu**[1], J. Haessler[1], M. Graff[2], J. Lundin[3], K. Jordahl[3], C. S. Carlson[4], E. Whitsel[5], L. A. Lange[6], A. Reiner[7], K. E. North[2], U. Peters[8], C. Kooperberg[9]; [1]Fred Hutch, Seattle, WA, [2]Univ North Carolina, Chapel Hill, NC, [3]Seattle, WA, [4]Fred Hutchinson Cancer Res Ctr, Seattle, WA, [5]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [6]Univ Colorado Denver, Aurora, CO, [7]Univ of Washington, Seattle, WA, [8]Fred Hutchinson CA Res Ctr, Seattle, WA, [9]Fred Hutchinson Cancer Res. Ctr., Seattle, WA
Disclosure Block: **Y. Hu:** None.
Circulating lipids are clinically associated with cardio-metabolic diseases. The phenotypic variance explained by identified genetic variants remains limited. DNA methylation has been linked to lipids in previous studies, and a total of 43 CpG sites mapped to 30 unique genes have been repeatedly linked to high-density lipoprotein cholesterol (HDL-c), low-density lipoprotein cholesterol (LDL-c), total cholesterol (TC), and triglycerides (TG) through methylome-wide association analyses. However, all these analyses were performed in European-ancestry populations with moderate sample sizes. In addition, knowledge regarding the joint potential effects of methylation patterns and non-genetic factors on lipids is extremely limited. In the Population Architecture Using Genomics and Epidemiology (PAGE) Study, a total of 7,864 participants from the Womenâ□□s Health Initiative (WHI) and the Atherosclerosis Risk in Communities (ARIC) studies with available HDL-c, LDL-c, TC, and TG concentrations were examined on the Illumina 450K array, mainly African (3,877) and European (3,079) Americans. After quality control, we performed methylome-wide association analysis on ~400,000 CpG sites and the four lipid traits with adjustment for age, sex, white blood cell proportions, technical covariates, and the first ten genetic principal components. We identified 252 novel CpG sites at 214 loci associated with at least one lipid trait (located more than 500kb away from any previously reported/known CpG sites, P

PrgmNr 3653 - Polygenic risk score analysis of congenital heart disease phenotypes: elucidating the contribution of common variants to congenital heart defects

View session detail

Author Block: **S. Spendlove**[1,2,3], L. Bondhus[2,4], J. Sul[5,1], V. A. Arboleda[5,1]; [1]InterDept.al Bioinformatics Program, UCLA, Los Angeles, CA, [2]Dept. of Pathology and Lab. Med., UCLA, Los Angeles, CA, [3]Psychiatry and Biobehavioral Sci., UCLA, Los Angeles, CA, [4]Dept. of Human Genetics, David Geffen Sch. of Med., UCLA, Los Angeles, CA, [5]UCLA, Los Angeles, CA
Disclosure Block: **S. Spendlove:** None.
Congenital heart disease (CHD) occurs in ~1% of live births. To date, the genetic architecture of CHD has shown that monogenic mutations are not completely penetrant and variably expressive, and that oligo- or poly-genic inheritance may be important. Polygenic risk scores (PRS) aggregate the effects of many common variants across the genome to investigate whether these variants as a group are associated with disease risk or severity. We hypothesized that PRS identify CHD cases that can be explained by the additive effects of common genetic variants and that PRS will be related to severity of the disease. Using genome-wide association study (GWAS) results from the UK Biobank (UKBB) as our base study, and whole genome sequencing data from the CHD cohort (n=711 trios) of the Gabriella Miller Kids First (GMKF) dataset as our target study, we developed a PRS for CHD. We used summary statistics from three CHD-related GWAS. GMKF whole-genome sequences underwent quality control and we narrowed our target data to 362 trios that aggregate with individuals of European genetic ancestry in principal component analysis. As the dataset is composed of trios, pseudo-controls were generated using untransmitted parental alleles. PRS were separately generated using both case-control and severity phenotypes. PRS were generated using all SNVs as well as using a subset of SNVs from top expressed fetal cardiac genes. The PRS from the â□□heart valve problem or heart murmurâ□□ GWAS gives significant results in European case-control PRS analyses using all SNVs (p=0.0081) as well as SNVs from genes expressed in fetal cardiac tissue (p=0.0082). The severity PRS using SNVs from genes expressed in fetal cardiac tissue was also significant (p=0.0077). These results can help explain the genetic basis of CHD of unknown genetic origin, as they indicate that common variants can contribute to risk and severity of CHD. In addition, by further elucidating the genetic basis of CHD this research lays the groundwork for new discoveries in treating CHD and its related health problems.

PrgmNr 3654 - Response to Polygenic Risk: Results of the MyGeneRank Mobile Application-Based Coronary Artery Disease Study

View session detail

Author Block: A. A. Torkamani[1], E. Muse[1], **S-F. Chen**[1], S. Liu[1], B. Fernandez[1], B. Schrader[1], B. Molparia[1], A. N. LeÃ³n[1], R. Lee[1], N. Pubbi[1], N. Mejia[1], C. Ren[2], A. El-kalliny[3], M. P. Ernesto[4], A. Hector[1], A. Ghoshal[1], R. Dias[1], D. Evans[1], K-Y. Chen[1], P. Zhang[1], N. E. Wineinger[5], E. Spencer[1], E. J. Topol[1]; [1]Scripps Res. Translational Inst., La Jolla, CA, [2]Stanford Univ., Palo Alto, CA, [3]DNA Visit, La Jolla, CA, [4]Centro de InvestigaciÃ³n y Asistencia en TecnologÃa y DiseÃ±o del Estado de Jalisco A.C., Guadalajara, Mexico, [5]Scripps Translational Sci. Inst., La Jolla, CA

Disclosure Block: **S. Chen:** None.

The degree to which polygenic risk scores (PRS) influence preventive health is the subject of debate, with few prospective studies completed to date. We developed a smartphone application for the prospective and automated generation, communication, and electronic capture of response to a PRS for coronary artery disease (CAD). We evaluated self-reported actions taken in response to personal CAD PRS information, with special interest in the initiation of lipid lowering therapy (NCT03277365). 19% (721 of 3,800) of valid participants provided complete responses for baseline and follow-up use of lipid lowering therapy. 20% (n = 19 of 95) of high genetic risk vs 7.9% (n = 8 of 101) of low genetic risk individuals initiated lipid lowering therapy at follow-up (p-value = 0.002). The initiation of both statin and non-statin lipid lowering therapy was associated with degree of genetic risk - 15.2% (n = 14 of 92) vs 6.0% (n = 6 of 100) for statins (p-value = 0.018) and 6.8% (n = 8 of 118) vs 1.6% (n = 2 of 123) for non-statins (p-value = 0.022) in high vs low genetic risk, respectively. Overall, degree of genetic risk was associated with use of any lipid lowering therapy at follow-up - 42.4% (n = 56 of 132) vs 28.5% (n = 37 of 130) (p-value = 0.009). We also find that CAD PRS information is perceived to be understandable, actionable, and does not induce health anxiety.

PrgmNr 3655 - Trans-ancestry PRSs of lipid traits to predict cardiovascular disease in the ancestrally diverse PAGE study

View session detail

Author Block: **M. Kim**[1], Z. Wang[2], Y. Cai[1], K. E. North[3], R. Smit[2], R. Loos[2], K. L. Young[3], S-A. M. Love[4], Z. Du[1], Y. Hu[1], J. Lundin[5], S. G. Buyske[6], S. Cullina[2], N. Abul-Husn[2], A. Thomas[7], L. Martin[8], S. E. Graham[9], C. J. Willer[10], C. Kooperberg[11], U. Peters[12], M. Graff[3], The Global Lipids Genetics Consortium, The PAGE Study [*MK and ZW contribute equally to this work]; [1]Fred Hutch, Seattle, WA, [2]Icahn Sch. of Med. at Mount Sinai, New York, NY, [3]Univ North Carolina, Chapel Hill, NC, [4]The Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [5]Seattle, WA, [6]Rutgers Univ, Piscataway, NJ, [7]UNC Chapel Hill, Chapel Hill, NC, [8]George Washington Univ., Washington DC, DC, [9]Univ MICHIGAN, Ann Arbor, MI, [10]Univ. of Michigan, Ann Arbor, MI, [11]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [12]Fred Hutchinson CA Res Ctr, Seattle, WA

Disclosure Block: **M. Kim:** None.

Recent studies have demonstrated the plausible extension of polygenic risk scores (PRS) to clinical settings for risk prediction and precision treatment. For the prediction of atherosclerotic cardiovascular disease (ASCVD), PRSs of known cardiovascular risk factors such as LDL, HDL, total cholesterol, and triglyceride levels can be used in addition to coronary artery disease (CAD) PRS. As most GWAS have an overrepresentation of populations of European ancestry, there is a concern about PRS generalizability across ancestries. Either ancestry-specific PRSs or trans-ancestral PRSs will be needed to predict diseases more accurately in underrepresented populations. The Global Lipids Genetics Consortium (GLGC) has developed trans-ancestral PRSs for lipid traits. Here, we apply the GLGC PRSs to the diverse Population Architecture through Genomics and Environment (PAGE) participants (17k African American, 3k Asian, 23k Hispanic, ~600 Native American, 2k Native Hawaiian, 23k European participants) for evaluation of LDL, HDL, total cholesterol, and triglyceride levels (R2~1.7% - 6%, adjusted for age, sex, self-reported race/ethnicity, use of lipid lowering medication, and 5 genetic PCs). When stratifying by ancestry, the explained variance is higher for European than for the average of the non-European populations (adjusted R2 European ~ 10%, 9.8%, 10%, 8.1%; average of the non-European~3.6%, 3.2%, 5.8%, 5.1% for HDL, LDL, total cholesterol, and triglyceride levels, respectively). The lowest correlations were observed for participants who self-identified as African American. We next examined the predictive accuracy of lipid PRSs for incident ASCVD 10-year risk in our ancestrally diverse population. Triglyceride and LDL PRSs outperformed the other two lipid PRSs and resulted in the discrimination for ASCVD C-statistic of 0.53. In comparison, non-genetic risk factors using the pooled cohort equation (PCE) score yielded an area under the receiver operating curve (AUC) of 0.71, which did not differ much by sex/race but decreased with increased age. Further, we evaluated the effects of adding lipid PRSs to the PCE score for ASCVD risk prediction. With added triglyceride and LDL PRS, the prediction performance of 10-year ASCVD risk improved among middle-aged participants (50-60 years old, increase in AUC 0.02) but not among older groups, shedding light on the potential clinical utility of lipid PRS. Overall, this study highlights the inconsistent performance of the trans-ancestral PRS for lipids traits across different ethnic groups, and pinpoints the subgroup populations that may benefit from the enhanced predictive ability of a PRS for CVD risk.

PrgmNr 3656 - Biometric trait sex heterogenous SNPs disproportionately influence predicted gene expression and diverse traits later in life

View session detail

Author Block: **M. Traglia**, M. Bout, L. A. Weiss; Univ. of California San Francisco, San Francisco, CA
Disclosure Block: **M. Traglia:** None.
Phenotypic differences across sexes are pervasive. We previously found that common SNPs differentially contributing to female and male secondary sexual characteristics are enriched in association signals in autism and other diseases [Mitra et al., 2016; Traglia et al, 2016]. We now leverage GIANT and UK Biobank cohorts to define a set of sexually dimorphic independent SNPs across 20 biometric phenotypes, and we investigated: 1) enriched biological processes, 2) disease/trait association signal, and 3) regulatory element overlap. Using a univariate approach combining female and male summary statistics of 20 anthropometric, metabolic and lung function traits, we extracted trait-specific sex-heterogenous SNPs (*Cochran test* P_Q -8). Then, we performed a multivariate meta-analysis of *Z-heterogeneity scores*. We extracted 2,320 independent *sex-heterogenous SNPs* within and across traits (MAF > 0.1%, *P* -8). For validation with empirical *P*-values (*EmpP*), we sampled 1,000 sets of 2,320 random *sex-homogeneous SNPs* matching the allele frequency and the combined-sex marginal effect for the leading sex-heterogeneous trait. *Sex-heterogenous SNPs* map in/near 1,325 genes (distance FDR and *EmpP* We investigated pleiotropic effects of sex-heterogenous SNPs, and we found association enrichment in 12 out of 18 diseases/traits at 5% alpha, ranging between 7.2% in ASD and late-onset asthma (*EmpP* 0.04) and 13.8% in educational attainment and schizophrenia (*EmpP* ex-heterogenous SNPs significantly overlap conserved regions across mammals, *TSS, promoter, coding*, *UTR 5â□□ regions* and introns (*EmpP* EmpP > 0.5). Interestingly, s*ex-heterogenous SNPs* strongly overlap with SNPs predicting gene expression, both across 49 GTEx tissues and across brain tissues only *(EmpP* . The excess of association signal for most tested diseases was retained compared to *eQTL-matched sex homogeneous SNPs* (*EmpP*

PrgmNr 3657 - Clinically impactful non-glycemic genetic effects on HbA1c yield missed detections of pre-diabetes, especially in African ancestry

View session detail

Author Block: **L. Stell**[1,2], h. Wang[3], I. Belitskaya-Levy[3], F. Wu[3], J. Lynch[4,5], K-M. Lee[4], D. R. Miller[6], S. Damrauer[7,8], T. Assimes[1,2], L. S. Phillips[9,10], M. Vujkovic[7,8], K-M. Chang[7,8], P. Tsao[1,2], VA Million Veteran Program, J. S. Lee[1,2]; [1]VA Palo Alto Hlth.Care System, Palo Alto, CA, [2]Stanford Univ. Sch. of Med., Stanford, CA, [3]VHA Cooperative Study Program Coordinating Ctr., Palo Alto, CA, [4]VA Salt Lake City Hlth.Care System, Salt Lake City, UT, [5]Univ. of Utah, Salt Lake City, UT, [6]VA Bedford Hlth.Care System, Bedford, MA, [7]Corporal Michael J. Crescenz VA Med. Ctr., Philadelphia, PA, [8]Univ. of Pennsylvania Perelman Sch. of Med., Philadelphia, PA, [9]Atlanta VA Med. Ctr., Atlanta, GA, [10]Emory Univ. Sch. of Med., Atlanta, GA

Disclosure Block: **L. Stell:** None.

Million Veteran Program (MVP) genetic and medical data from a large, diverse population are ideal for addressing unresolved questions about non-glycemic effects on HbA1c. We investigated variants associated with pre-diabetes and whether non-glycemic effects on HbA1c might affect detection of glycemic dysregulation in health care settings.

To identify candidate genes, we conducted a GWAS of patients with pre-diabetes versus normoglycemic controls with status determined by electronic health ICD codes, lab results and pharmacy data. A sample of 128,098 European (EUR), 30,726 African (AFR) and 12,672 Hispanic (AMR) ancestries, analyzed separately, identified 44 independent significant (PAfter tripling sample sizes by adding Type 2 diabetes (T2D) patients and recent enrollees, erythrocytic GRS (eGRS) explains 4% of HbA1c variance in AFR but Thus, non-glycemic genetic effects can lower HbA1c level enough to lead to missed detection of pre-diabetes, especially in patients with African ancestry.

PrgmNr 3658 - Genome-wide association study of sphingolipid traits in diabetic patients reveals distinct cis-regulatory regions for sphingolipid pathway-associated genes

View session detail

Author Block: **C. Simeone**[1,2], S. G. Frodsham[2], A. M. Poss[3], R. J. Nicholson[3], M. H. Pezzolesi[2], W. L. Holland[3], S. A. Summers[3,4], M. G. Pezzolesi[1,2,3]; [1]Dept. of Human Genetics, Univ. of Utah, Salt Lake City, UT, [2]Div. of Nephrology and Hypertension, Univ. of Utah, Salt Lake City, UT, [3]Dept. of Nutrition and Integrative Physiology, Univ. of Utah, Salt Lake City, UT, [4]Sch. of Med., Univ. of Utah, Salt Lake City, UT

Disclosure Block: **C. Simeone:** None.

Background Sphingolipids are bioactive molecules responsible for cellular responses such as cell growth, death, and membrane dynamics. Recently, ceramides, a class of sphingolipids, have gained attention as antagonists for insulin signaling and drivers of metabolic diseases, including diabetes. While most research has focused on manipulating sphingolipid pathway genes to alter ceramide levels, little is known about the genetic determinants of their regulation. Here, we investigated genetic loci associated with circulating sphingolipid levels and identified variants located in cis-regulatory regions of key members of sphingolipid biosynthetic pathways. **Methods** We performed a genome-wide association study (GWAS) on 39 serum lipid traits, including dihydroceramides and ceramides, in 293 patients from the Utah Diabetes and Diabetic Complications Study. To identify variants driving cis-regulation of sphingolipid levels, we focused our analyses on single nucleotide polymorphisms (SNPs) that i) are reported as expression quantitative trait loci (eQTLs) from the Genotype-Tissue Expression (GTEx) Project for sphingolipid pathway genes, and ii) localize to candidate cis-regulatory elements from the Encyclopedia of DNA Elements (ENCODE) Project. **Results** Top associations were identified at several novel candidate genes, including *COL4A2*, *CPPED1*, and *THSD7B,* that may influence ceramide levels. Focusing on eQTL-associated SNPs for sphingolipid pathway genes, we identified rs35237501, near *CERS4*, in a cis-regulatory element associated with Cer16:0 levels. Additionally, we identified an intronic SNP (rs11006155) in *SGMS1,* associated with Cer16:0 levels. Additional associations were identified in cis-regulatory regions near *SPTLC2* for both Cer16:0 and Cer18:0. **Conclusion** Our study has identified genetic variants that are associated with both circulating sphingolipid levels, including lipotoxic ceramides, and expression of genes involved in the sphingolipid pathway. Importantly, these variants may predispose individuals to aberrant levels of ceramides, which can further influence disease progression in metabolic disorders. Further examination into these variants is necessary to understand the extent of their influence on ceramide levels and their role in disease.

PrgmNr 3659 - Grip strength, adiposity, and genetic risk of non-alcoholic fatty liver disease: Findings from the UK Biobank study

View session detail

Author Block: **T. M. Schnurr**[1], S. Figueroa Katz[1], J. M. Justesen[1], J. W. O'Sullivan[2], P. Saliba-Gustafsson[1], T. L. Assimes[3], I. Carcamo-Orive[4], E. Ashley[5], T. Hansen[6], J. W. Knowles[1]; [1]Dept. of Med., Div. of Cardiovascular Med. and Cardiovascular Inst., Stanford Univ., Stanford, CA, [2]Dept. of Med., Div. of Cardiology, Stanford Univ. Sch. of Med., Stanford, CA, [3]VA Palo Alto Hlth.Care System, Palo Alto CA, USA & Dept. of Med., Stanford Univ. Sch. of Med., Stanford, CA, [4]Dept. of Med., Div. of Cardiovascular Med. and Cardiovascular Inst. & Stanford Diabetes Res. Ctr., Stanford Univ., Stanford, CA, [5]Div. of Cardiology, Dept. of Med., Stanford Univ. Sch. of Med. & Dept. of BioMed. Data Sci., Stanford Univ., Stanford, CA, [6]Novo Nordisk Fndn. Ctr. for Basic Metabolic Res., Univ. of Copenhagen, Copenhagen, Denmark

Disclosure Block: **T.M. Schnurr:** None.

Background and Aims: Non-alcoholic fatty liver disease (NAFLD) is a major cause of morbidity and mortality with an estimated global prevalence of 25%. Obesity and a sedentary lifestyle are risk factors for NAFLD. We investigated whether the genetic predisposition to NAFLD is altered by muscular fitness (as assessed by hand grip strength) and adiposity. We hypothesized that grip strength attenuates, while adiposity accentuates, the genetic risk of NAFLD.

Methods: Elevated levels of the liver enzyme alanine aminotransferase (ALT) in the absence of excessive alcohol intake and liver-related diseases are routinely used to identify individuals with NAFLD. Using UK Biobank data, we excluded individuals with excessive alcohol intake and liver-related diseases and restricted our analysis to the White British ancestry subset with available genotype information (N= 242,524). Genetic predisposition for NAFLD was quantified using a genetic risk score (GRS) comprising 77 loci that are associated with chronically elevated ALT levels in the Million Veteran Program. We used relative instead of absolute grip strength to reduce confounding effects of body size. Relative grip strength was calculated as the average of right- and left-hand measurements divided by whole body fat-free mass. We used linear regression models to test for associations of grip strength, body-mass index (BMI), and the GRS with ALT levels. We also tested for GRS*grip strength and GRS*BMI interactions on ALT. All analyses were adjusted for age, gender, Townsend deprivation index, and UK Biobank assessment center. Analyses including the GRS were additionally adjusted for genotyping array and the first ten genome-wide principal components. Phenotypic traits were rank inverse normally transformed to approximate normal distribution, and the effect sizes are thus reported in standard deviation (SD) units of the inverse normally transformed trait.

Results: Higher grip strength is associated with decreased ALT levels (Î²=-0.090 SD, p-16), while higher BMI is associated with increased ALT levels (Î²=0.28 SD, p-16). A higher GRS is associated with increased ALT levels (Î²=0.045 SD/allele, p-16), but grip strength attenuates this effect (p=1.1Ã10^{-5} for GRS-grip strength interaction), while adiposity accentuated this effect (p-16 for GRS-BMI interaction).

Conclusions: We confirm previous reports that adiposity amplifies the genetic risk of NAFLD, and discover that hand grip strength attenuates the genetic risk of NAFLD. Our results highlight the importance of normal weight and muscular fitness in the context of genetic predisposition to NAFLD.

PrgmNr 3660 - Integration of transcriptome genetics, metabolic perturbations, and trait associations identifies high-confidence cardiometabolic risk genes

View session detail

Author Block: **M. J. Gloudemans**[1,2], B. Balliu[3], D. Nachun[4,5], M. G. Durrant[4], E. Ingelsson[6], W. Martin[7], T. Quertermous[6,8], S. B. Montgomery[2,4], J. W. Knowles[6,8,9], I. Carcamo-Orive[6,8]; [1]Bioinformatics Training Program, Stanford Univ., Stanford, CA, [2]Dept. of Pathology, Stanford Univ., Stanford, CA, [3]Dept. of Computational Med., UCLA, Los Angeles, CA, [4]Dept. of Genetics, Stanford Univ., Stanford, CA, [5]Dept. of Immunology, Stanford Univ., Stanford, CA, [6]Dept. of Med., Div. of Cardiovascular Med. and Cardiovascular Inst., Stanford Univ., Stanford, CA, [7]Dept. of Pediatrics and Adolescent Med., Div. of Pediatric Endocrinology, Ulm Univ., Ulm, Germany, [8]Diabetes Res. Ctr., Stanford Univ., Stanford, CA, [9]Prevention Res. Ctr., Stanford Univ., Stanford, CA

Disclosure Block: **M.J. Gloudemans:** None.

Identification of causal genes for GWAS loci of cardiometabolic traits has been extremely challenging, and our understanding of how physiological and pharmacological stimuli modulate genetic risk for these loci is limited. Moreover, insulin resistance (IR), a common feature of type 2 diabetes, obesity and dyslipidemia, lacks well-powered GWAS and therefore few causal genes have been identified for IR. Here, we perform LD-adjusted integrative colocalization analyses across nine cardiometabolic traits combined with eQTLs and sQTLs from five metabolically relevant human tissues (subcutaneous and visceral adipose, skeletal muscle, liver, and pancreas). We identify 470 colocalized loci and further prioritize 207 loci with a single colocalized gene. To elucidate upstream regulators and functional mechanisms for these genes, we measure their transcriptional responses to 21 physiological and pharmacological cardiometabolic regulators in human adipocytes, hepatocytes, and skeletal muscle cells, and map their protein-protein interactions. The combination of genetic associations from our state-of-the-art colocalization approach with transcriptional responses to a large panel of metabolic perturbations provides a knowledge-based list of likely causal genes and potential upstream regulators in the context of IR-associated cardiometabolic risk.

PrgmNr 3661 - Investigation of Rapid-onset Obesity with Hypothalamic dysfunction, Hypoventilation and Autonomic Dysregulation (ROHHAD) by comprehensive whole genome sequencing

View session detail

Author Block: **S. Barclay**[1], A. Crawford[2], R. Wang[2], C. Rand[3], R. J. Taft[2], K. Kurek[1], N. T. Bech-Hansen[1], D. E. Weese-Mayer[3]; [1]Alberta Childrenâ□□s Hosp. Res. Inst. and Cumming Sch. of Med., Univ. of Calgary, Calgary, AB, Canada, [2]Illumina Inc., San Diego, CA, [3]Ann & Robert H. Lurie Childrenâ□□s Hosp. of Chicago and Stanley Manne Children's Res. Inst. Northwestern Univ. Feinberg Sch. of Med., Chicago, IL

Disclosure Block: **S. Barclay:** None.

Rapid-onset Obesity with Hypothalamic dysfunction, Hypoventilation and Autonomic Dysregulation (ROHHAD) is an ultra-rare pediatric disorder without a clear-cut etiology. Less than 200 cases have been described globally. The hallmark characteristic of affected children is rapid weight gain of 20-30 pounds over the course of 3-6 months before the age of 10, in addition to pronounced hypoventilation and autonomic dysregulation. Identification of the genetic etiology of ROHHAD would provide a much-needed diagnostic tool and insight into its pathogenesis and might offer intervention strategies. Prior candidate gene and whole exome sequencing studies in modestly sized cohorts have failed to discover a genetic basis for ROHHAD to date. Here, we undertook whole genome sequencing of the Lurie Childrenâ□□s Hospital cohort of 40 ROHHAD families (6 quads, 3 trios, 2 duos and 29 singletons), using DNA from peripheral blood or saliva samples, in an effort to establish a genetic basis for ROHHAD. In our extensive clinical study of ROHHAD, we have not identified affected siblings or parents, suggesting *de novo* mutations as a potential disease mechanism. In addition, ROHHAD has been described in 2 sets of identical twins discordant for the ROHHAD phenotype, supporting this theory, and possibly extending it to somatic, as opposed to germline, *de novo* variants. Despite an investigation that spanned all known disease genes, included SNVs, indels, SVs, repeat expansions and CNVs, and prioritized all potentially disease-causing variants based on both heuristic and AI-prioritization, no candidate disease genes were identified among ROHHAD probands. We also examined the coding and intronic regions of previously suggested ROHHAD candidate genes but found no consistent association in our dataset. Taken together, these results suggest that the underlying pathophysiology of ROHHAD is unlikely to stem from simple Mendelian inheritance. While we have previously reported a similar conclusion from exome sequencing in a subset of this cohort, we have now extended the analysis to include non-coding genetic variants that may affect splicing or gene regulation, structural variants, and repeat expansions--all of which cannot be reliably detected by exome sequencing. Now that the whole genomes of the expanded ROHHAD cohort (n=40) have been thoroughly searched for â□□simpleâ□□ genetic explanations, the ROHHAD research community can confidently move forward in exploring other tissue sources and other etiologies including complex/multifactorial genetic inheritance, somatic genetic variants, autoimmune pathology, a paraneoplastic syndrome, and an epigenetic basis.

PrgmNr 3662 - Trans-ancestry meta-analysis of the X chromosome for obesity-related traits in over 600K individuals

View session detail

Author Block: **J. Arias**[1], E. Marouli[2], S. Vedantam[3,4], X. Yin[5], E. P. Wilson[6], L. Yengo[7], M. Graff[8], K. L. Young[8], A. E. Locke[9,10], C. Lindgren[11,12], J. N. Hirschhorn[3,13], K. L. Mohlke[6], R. Loos[14], A. Justice[15], S. I. Berndt[16]; [1]Div. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD, [2]William Harvey Res. Inst., Barts and The London Sch. of Med. and Dentistry, Queen Mary Univ. of London, London, United Kingdom, [3]Dept. of Endocrinology, Boston Children's Hosp., Boston, MA, [4]Broad Inst. of MIT and Harvard, Cambridge, CA, [5]Dept. of Biostatistics and Ctr. for Statistical Genetics, Univ. of Michigan, Ann Arbor, MI, [6]Dept. of Genetics, Univ. of North Carolina, Chapel Hill, Chapel Hill, NC, [7]Program in Complex Trait Genomics, Univ. of Queensland, St Lucia, Brisbane, Australia, [8]Dept. of Epidemiology, Univ. of North Carolina, Chapel Hill, NC, [9]McDonnell Genome Inst., Washington Univ. in St. Louis, St. Louis, MO, [10]Dept. of Med., Washington Univ. Sch. of Med., St. Louis, MO, [11]Big Data Inst., Univ. of Oxford, Oxford, United Kingdom, [12]Wellcome Trust Ctr. for Human Genetics, Univ. of Oxford, Oxford, United Kingdom, [13]Dept.s of Pediatrics and Genetics, Harvard Med. Sch., Boston, MA, [14]The Icahn Sch. of Med. at Mount Sinai, New York, NY, [15]Dept. of Population Hlth.Sci., Geisinger Hlth.System, Danville, PA, [16]Dept. of Cancer Epidemiology and Genetics, Natl. Cancer Inst., Rockville, MD

Disclosure Block: **J. Arias:** None.

Sexual dimorphism has been observed for autosomal variants in genetic association studies of fat distribution, but less so for overall obesity. Although the X chromosome represents ~5% of the genome, few genome-wide association studies (GWASs) have investigated the contribution of the X chromosome to obesity. To discover new loci and gain a better understanding of the role that the X chromosome plays in the genetic architecture of obesity, we conducted trans-ancestry meta-analyses of GWASs of body mass index (BMI) and waist-to-hip ratio unadjusted for BMI (WHR). Meta-analyses included up to 655,573 individuals (307,149 males and 348,424 females) from five major ancestries (African American, East Asian, European, Hispanic American, and South Asian) and were imputed to the 1000 Genomes Phase 3 and Haplotype Reference Consortium reference panels. Each study was analyzed separately for men and women using a linear mixed model. After applying quality control measures, the results were meta-analyzed using a fixed-effects model. In the trans-ancestry meta-analysis for BMI, we identified 9 loci at a stringent genome-wide significance level of P 1 million individuals with plans to conduct further fine-mapping of identified loci within and across ancestries and bioinformatic analyses to gain further insight into the biology of these variants.

PrgmNr 3663 - Ancestry-aware polygenic risk scores identify host factors behind susceptibility and disease outcomes associated to SARS-CoV-2 infection from a nationwide longitudinal direct-to-consumer survey of 30,000 individuals

View session detail

Author Block: **M. Vernekar**[1,2], A. Rojas-MuÃ±oz[3], A. Lopez Pineda[4], K. Numakura[1,2], Y. Matsuda[1,2], N. Katsanis[3], T. Takano[1,2], C. D. Bustamante[5,3,6,7]; [1]Awakens Inc., Berkeley, CA, [2]Awakens K.K., Tokyo, Japan, [3]Galatea Bio Inc., Hialeah, FL, [4]Amphora Hlth., Morelia, Mexico, [5]Stanford Univ. Sch. of Med., Stanford, CA, [6]Univ. of Miami Sch. of Business, Miami, FL, [7]Chan Zuckerberg Biohub, San Francisco, CA
Disclosure Block: **M. Vernekar:** None.
Background: Despite advances in genetic medicine, our understanding of pleiotropy and variable expressivity remains poor. This persistent knowledge gap represents a key impediment in both stratifying health care delivery and predicting risk across a range of disorders. SARS-CoV-2 has highlighted this phenomenon by virtue of the fact that infected individuals present acutely varied manifestations, ranging from asymptomatic to severe complications or even death. Although some of this variability can be attributed to the genetic variation in the viral genome, underlying host conditions, and health disparities, recent data also suggest a significant contribution of the host genome to SARS-CoV-2 sequelae. The pandemic, together with our technological capacity to systematically test for infection and to genotype large numbers of individuals offers an unprecedented opportunity to identify host factors associated with disease susceptibility and progression across different ancestries. **Methods:** We collected information regarding COVID-19 risk factors and host genetics via an online survey from research participants belonging to a large nationwide direct-to-consumer genetic testing research cohort in the USA. Age, sex, ancestry, and preexisting conditions were explored as risk factors using bivariate and multivariate logistic regression models. **Results:** From June to November 2020, we collected information from 30,169 participants. From 2,851 individuals tested for COVID-19, we identified 220 (7%) who reported a positive COVID-19 test, of which 29 (1%) reported hospitalization with COVID-19 symptoms. Of those, 21 individuals required respiratory support in the ICU. Our logistic regression models predicted that non-European ancestry and preexisting health conditions were associated with a higher risk of hospitalization with an accuracy of 0.76 and 0.92 respectively. When controlled by cofactors, non-Europeans over 45 years and participants managing high blood pressure exhibited the highest risk (OR 2.44 and 1.54 respectively). Our polygenic risk score (PRS) analysis of phenotypes related to COVID-19 diagnosis and severity will likely improve our ability to predict genetic risk and help us locate specific loci contributing to this score. **Conclusions:** Preexisting genetic cohorts offer an unprecedented opportunity to identify genetic correlates of disease onset and progression. Our ongoing online survey is poised to identify multiple genetic determinants of disease susceptibility and severity for COVID-19 and, together with our local ancestry inference algorithm, help establish how these variants differ between ancestry groups.

PrgmNr 3665 - Common and rare variant analyses combined with single-cell multiomics reveal cell-type-specific molecular mechanisms of COVID-19 severity

View session detail

Author Block: **S. Zhang**[1], J. Cooper-Knock[2], A. Weimer[1], C. Harvey[2], T. Julian[2], C. Wang[3], J. Li[3], S. Furini[4], E. Frullanti[4], F. Fava[4], A. Renieri[4], C. Pan[5], J. Song[1], P. Billing-Ross[1], P. Gao[1], X. Shen[1], I. Timpanaro[6], K. Kenna[6], VA Million Veteran Program, GEN-COVID Network, M. Davis[1], P. TSAO[7], M. P. Snyder[8]; [1]Stanford Univ. Sch. of Med., Palo Alto, CA, [2]Univ. of Sheffield, Sheffield, United Kingdom, [3]Univ. of California, San Francisco, San Francisco, CA, [4]Univ. of Siena, Siena, Italy, [5]Stanford Univ, Palo Alto, CA, [6]Univ. Med. Ctr. Utrecht, Utrecht, Netherlands, [7]VAPAHCS, Palo Alto, CA, [8]Stanford Univ, Stanford, CA
Disclosure Block: **S. Zhang:** None.
The determinants of severe COVID-19 in non-elderly adults are poorly understood, which limits opportunities for early intervention and treatment. Here we present novel machine learning frameworks for identifying common and rare disease-associated genetic variation, which outperform conventional approaches. By integrating single-cell multiomics profiling of human lungs to link genetic signals to cell-type-specific functions, we have discovered and validated over 1,000 risk genes underlying severe COVID-19 across 19 cell types. Identified risk genes are overexpressed in healthy lungs but relatively downregulated in severely diseased lungs. Genetic risk for severe COVID-19, within both common and rare variants, is particularly enriched in natural killer (NK) cells, which places these immune cells upstream in the pathogenesis of severe disease. Mendelian randomization indicates that failed NKG2D-mediated activation of NK cells leads to critical illness. Network analysis further links multiple pathways associated with NK cell activation, including type-I-interferon-mediated signalling, to severe COVID-19. Our rare variant model, PULSE, enables sensitive prediction of severe disease in non-elderly patients based on whole-exome sequencing; individualized predictions are accurate independent of age and sex, and are consistent across multiple populations and cohorts. Risk stratification based on exome sequencing has the potential to facilitate post-exposure prophylaxis in at-risk individuals, potentially based around augmentation of NK cell function. Overall, our study characterizes a comprehensive genetic landscape of COVID-19 severity and provides novel insights into the molecular mechanisms of severe disease, leading to new therapeutic targets and sensitive detection of at-risk individuals.

PrgmNr 3666 - Genome-wide association analyses of common infections in a large practice-based biobank

View session detail

Author Block: **L. Jiang**[1], C. M. Shaffer[1], A. Dickson[2], M. Ormseth[1,2], V. Kerchberger[3], L. Daniel[1,2], B. Carranza leon[4], N. J. Cox[5], C. P. Chung[1,2], W-Q. Wei[6], M. Stein[1,7], Q. FENG[3]; [1]Div. of Clinical Pharmacology, Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [2]Div. of Rheumatology and Immunology, Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Vanderbilt Univ. Med. Ctr., Nashville, TN, [4]Div. of Diabetes, Endocrinology and Metabolism, Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [5]Vanderbilt Genetic Inst., Dept. of Med., Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Dept. of BioMed. Informatics, Vanderbilt Univ. Med. Ctr., Nashville, TN, [7]Dept. of Pharmacology, Vanderbilt Univ., Nashville, TN

Disclosure Block: **L. Jiang:** None.

Introduction: Infectious diseases are one of the most common causes of morbidity and mortality worldwide. Susceptibility to infection is highly heritable; however, little has been done to identify the genetic determinants underlying common infectious diseases. Previous associations were sought in small candidate gene studies, and there have been few GWAS applicable to US patients. The one relevant GWAS was performed using 23andMe data, but only self-reported infections.

Methods: We leveraged the EHR-based biobank at Vanderbilt and physician-confirmed diagnosis codes (International Classification of Disease [ICD]) to identify cases for 12 infectious diseases studied in the 23andMe GWAS: urinary tract infection (UTI), pneumonia, chronic sinus infections, otitis, candidiasis, hepatitis C, streptococcal pharyngitis, herpes zoster, cold sores, hepatitis B, infectious mononucleosis, and tuberculosis (TB). The number of cases ranged from 9359 (UTI) to 102 (TB). We selected controls from patients with no ICD code for the candidate disease and matched by year of birth, gender, and calendar year at 1st and last VUMC visits. We conducted GWAS using SAIGE (covariates: age, gender, EHR length, and PCs) and transcriptome-wide analysis (TWAS) using S-PrediXcan and Genotype Tissue Expression.

Results: We replicated three 23andMe hits (p-3); rs2808290-C (OR, 1.09; 95% CI [1.02-1.16]; $p=9.6\tilde{A}\square10^{-3}$) and rs114947103-C (OR, 1.09; 95% CI [1.00-1.18]; p=0.04) associated with ear infections. We also identified 3 novel associated regions (p-8). Top hits included: rs113235453-G for otitis (OR, 1.5; 95% CI [1.3-1.7]; $p=3.04\tilde{A}\square10^{-8}$); rs10422015-T for candidiasis (OR, 1.3; 95% CI [1.2-1.4]; $p=3.11\tilde{A}\square10^{-8}$); and rs1242993-C for TB (OR, 2.5; 95% CI [1.8-3.4]; $p=5.03\tilde{A}\square10^{-8}$). In TWAS, we identified 4 gene-disease associations: SLC30A9 associated with otitis ($p=8.06\tilde{A}\square10^{-7}$); LRP3 ($p=3.91\tilde{A}\square10^{-7}$) and WDR88 associated with candidiasis ($p=1.95\tilde{A}\square10^{-6}$); and AAMDC associated with hepatitis B ($p=1.51\tilde{A}\square10^{-6}$).

Conclusion: We conducted GWAS and TWAS for 12 common infectious diseases and identified novel genetic contributors to the susceptibility of infection diseases.

PrgmNr 3667 - Mid-density Gene Expression Profiling of SARS-COV-2 infected samples using Applied Biosystems TaqMan Flexible Array Panels

View session detail

Author Block: **A. Gupta**, S. Jackson; Thermo Fisher Scientific, South San Francisco, CA
Disclosure Block: **A. Gupta:** Salary/Employment; Thermo Fisher Scientific.
The broad spectrum of clinical manifestations from SARS-COV-2 infection, as well as the observed risk factors for severe disease, highlight the importance of understanding molecular mechanisms underlying COVID-19 disease development and progression. Research studies have identified a large number of host proteins that play roles in viral entry, innate immune response, or immune signaling during infection. The ability to interrogate subsets of these genes simultaneously within SARS-COV-2 infected samples is critical to understanding how their expression contribute to phenotypic variability of COVID-19 disease. To bridge this gap, we will use flexible TaqMan array panels designed by Applied Biosystems specifically for targeting the most cited genes related to entry and restriction factors as well as cytokines, chemokines, and growth factors. Each array features a curated list of predesigned TaqMan Gene Expression Assays that can be modified according to research objectives. In this initial study, these arrays will be used to highlight gene expression patterns that exist within confirmed SARS-COV-2 positive and negative nasopharyngeal swab samples and demonstrate the utility of these panels for gene expression profiling of SARS-COV-2 infected samples at medium throughput and scale.

PrgmNr 3668 - Multi-ancestry GWAS reveals association between absolute neutrophil count and *SETD1B*

View session detail

Author Block: **E. A. Rosenthal**[1], I. B. Stanaway[2], T. Li[2], D. R. Crosslin[3], D. S. Carrell[4], E. B. Larson[4], J. J. Connolly[5], B. Namjou[6], K. Kiryluk[7], J. C. Denny[8], E. W. Karlson[9], T. E. Kitchner[10], G. P. Jarvik[11]; [1]Univ. of Washington Sch. of Med., Seattle, WA, [2]Univ. of Washington, Seattle, WA, [3]Tulane Univ. Sch. of Med., New Orleans, LA, [4]Kaiser Permanente, Seattle, WA, [5]Children's Hosp. of Philadelphia, Philadelphia, PA, [6]Children's Hosp. Med. Ctr., Cincinnati, OH, [7]Columbia Univ. - Coll. of Physicians and Surgeons, New York, NY, [8]NIH, Kensington, MD, [9]Harvard Med. Sch., Boston, MA, [10]Marshfield Clinic Res. Inst., Marshfield, WI, [11]Univ Washington Med Ctr., Seattle, WA

Disclosure Block: **E.A. Rosenthal:** None.

Neutrophils are key components of immune response and make up the largest portion of white blood cells. Understanding the biology of neutrophils may reveal insights into interindividual variation in immune response and susceptibility to infection. We performed GWAS on absolute neutrophil count (ANC) on participants from ten sites from the Electronic Medical Records and Genomics (eMERGE) study, adjusting for known associated loci to detect novel signals. The data set include N=37,199 unrelated participants from multiple ancestries (N= 35,306 European ancestry, N=1,256 African ancestry, N=429 Asian/Pacific Islander/Native American ancestry; self-reported) with phenotype and genotype data. Previously, genotypes from multiple arrays were aligned and then used to impute genome-wide using the Michigan Imputation Server (PMID: 30298529). After quality control, there were approximately 10M SNPs in HWE with MAF > 0.005 and imputation r^2 > 0.3 for this study. As multiple records at multiple time points were available, we used the square root mean ANC and adjusted it for mean age, sex, the first four principal components of ancestry as well as independent (r^2 ACKR1) on chromosome 1 (2 SNPs, F-test p=3E-95), HLA region on chromosome 6 (4 SNPs, F-test p=5E-5), a known eQTL, rs542000 and *CDK6* on chromosome 7 (p=1.1E-8 and 8.8E-10, respectively) and *PSMD3* on chromosome 17 (5 SNPs, F-test p=3E-49). Using a p-value cutoff of 5E-8 and adjusting for inflation factor (Î»=1.02), the resulting GWAS detected signals at rs34599082 (p=1.9E-8), a missense in exon 2 of *ACKR1*, which is common in European and South Asian ancestries (MAF=0.01), as well as other previously reported loci including *CSF3R* on chromosome 1 (rs3917932, p=4.3E-11), an intergenic region on chr3 (rs6782812, p=1.4E-8) and a locus previously reported in Asian ancestry populations on chr19 (rs144284241, p=1.3E-10). Interestingly, we found a potentially novel locus on chromosome 12 with moderate evidence of association (rs117724148, p=2.1E-8). This SNP is an intronic variant of *SETD1B*, which plays a role in trimethylation of histones and is reported to regulate hematopoiesis in mice (PMID: 29916805).

PrgmNr 3669 - Predicting host response to Covid-19

View session detail

Author Block: **R. Allman**[1], J. McDevitt[2], N. M. Murphy[1], G. S. Dite[1]; [1]Genetic Technologies Limited, Fitzroy, Australia, [2]Inifinity BiologiX LLC, Piscataway, NJ
Disclosure Block: **R. Allman:** Salary/Employment; Genetic Technologies Limited. Receipt of Intellectual Property Rights/Patent Holder; Genetic Technologies Limited.
Epidemiological analysis of clinical risk factors associated with disease severity in people with COVID-19 began appearing on pre-print sites and in the literature soon after the declaration of COVID-19 as a pandemic. The availability of genetic data associated with disease severity has been slower, despite the establishment of the COVID-19 Host Genetics Initiative. Identification of host genetic factors that predispose individuals to severe COVID-19 is important, not only for understanding the disease and guiding the development of treatments, but also for risk prediction. Using SARS-CoV-2 positive participants from the UK Biobank, we developed and validated a clinical and genetic model to predict risk of severe COVID-19. We used source of test result as a proxy for severity of disease. Inpatient results were considered cases and outpatient results were considered controls. Of the 7621 eligible participants, 2205 (28.9%) were cases and 5416 (71.7%) were controls. We used multivariable logistic regression on a 70% training dataset and used the remaining 30% for validation. The resultant model comprised age, sex, ethnicity, body mass index, haematological cancer, non-haematological cancer, respiratory disease (excluding asthma), hypertension, diabetes, kidney disease, cerebrovascular disease, and 7 single-nucleotide polymorphisms (SNPs), with each SNP being incorporated as an independent risk factor.
The model was strongly associated with severe COVID-19 (odds ratio per quintile of risk=1.77, 95% confidence interval [CI]=1.64, 1.90). Discrimination (as assessed by the area under the receiver operating characteristic curve) was 0.732, 95% CI=0.708, 0.756. Calibration was assessed using logistic regression of the log odds of the risk score, with the model showing no evidence of over- or under-estimation of risk (Î±=â□□0.08; 95% CI=â□□0.21, 0.05) and no evidence or over-or under-dispersion of risk (Î²=0.90, 95% CI=0.80, 1.00).
We also attempted to validate the SNP panels identified by Pairo-Castinera et al. and found that when combined as polygenic risk scores, their two SNP panels had no ability to distinguish severe COVID-19 from non-severe COVID-19. We suspect that our inability to replicate their findings is because they used population controls instead of non-severe cases of COVID-19 and have identified SNPs associated with propensity to become infected rather than SNPs associated with severe COVID-19. Relying on population controls may not just attenuate results as Pairo-Castinera et al. claim. We propose that results from studies that use population controls to study severe COVID-19 should be interpreted with care.

PrgmNr 3670 - Identification of Gene Expression Patterns Associated With Polycystic Ovary Syndrome

View session detail

Author Block: **S. Lyle**, B. Drogemoller; Univ. of Manitoba, Winnipeg, MB, Canada
Disclosure Block: **S. Lyle:** None.
Background: Polycystic ovary syndrome (PCOS) is a common endocrine disorder, which is accompanied by a variety of comorbidities including metabolic, reproductive, and psychiatric disorders. PCOS is significantly influenced by genetics both in terms of germline genetic variants and epigenetic influences. While there are a number of treatment options for patients with PCOS, the administration of these therapeutics is done using a trial-and-error-based approach, which fails to provide optimal treatment outcomes for many patients.**Methods:** We performed a transcriptome-wide association study (TWAS) to uncover heritable gene expression profiles that are associated with PCOS. After colocalization analyses, the lead variants associated with both gene expression and PCOS were included in a Phenome-wide association study (PheWAS) using the UK Biobank (UKBB) and Finngen data. Drug repurposing using the data housed in CMap were performed to identify small molecules with significant similarity and dissimilarity to the PCOS gene expression profiles identified via the TWAS.**Results**: The TWAS analyses revealed that increased expression of *ARL14EP* was significantly associated with PCOS (P=1.6x10^{-6}), with increased expression of this gene shown to be of particular relevance to the female reproductive organs i.e. ovaries, uterus and vagina. Upon colocalization evaluation, rs4071559 was shown to be associated with both an increase in PCOS risk and *ARL14EP* expression. PheWAS analyses revealed that this variant was associated with a number of traits of relevance to PCOS, including increased length of menstrual cycles (P=8.5x10^{-34}), which is a key clinical feature of PCOS. The CMap analysis revealed a number of possible therapeutic candidates, including prednisone, which induces ovulation in patients with PCOS by directly reducing adrenal androgen production.**Discussion**: This TWAS of PCOS, being the first of its kind, has provided evidence for the role of *ARL14EP* in PCOS disease mechanisms. The drug repurposing analyses have opened avenues for the exploration of redirecting therapeutics to target the various physiology ailments associated with PCOS. By uncovering a genetic candidate of PCOS, this study has contributed to generating knowledge that can be used to guide strategies to improve the efficiency, accuracy, timing and safety of therapeutics used in the treatment of PCOS.

PrgmNr 3671 - Mendelian randomization analyses reveal no evidence of causal relationships between chronic health conditions and sporadic and recurrent miscarriage

View session detail

Author Block: **J. Painter**[1,2,3], T. Laisk[4], C. Lindgren[5], S. E. Medland[6,2,3]; [1]QIMR Berghofer Med. Res. Inst., Herston, Australia, [2]Univ. of Queensland, Brisbane, Australia, [3]Queensland Univ. of Technology, Brisbane, Australia, [4]Inst. of Genomics, Univ. of Tartu, Tartu, Estonia, [5]Big Data Inst., Oxford Univ., Oxford, Oxfordshire, United Kingdom, [6]QIMR Berghofer, Brisbane, Australia
Disclosure Block: **J. Painter:** None.
Miscarriage is estimated to end between 10-25% of clinically confirmed pregnancies. Observational studies have suggested that risk of miscarriage is increased by numerous chronic health conditions in mothers, such as uncontrolled type 2 diabetes, endometriosis, polycystic ovarian syndrome, hyper- and hypothyroidism, and various inflammatory conditions. However, definitive causal relationships between these potential risk factors and either sporadic or recurrent miscarriage have not yet been demonstrated. Using summary results from large-scale genome-wide association studies (GWAS) we have conducted two-sample Mendelian randomization association and sensitivity analyses to investigate relationships between miscarriage and chronic health conditions. Instrumental variables were constructed using 6-46 genetic variants significantly associated with type 2 diabetes, fasting glucose, fasting insulin, endometriosis, polycystic ovarian syndrome, FT4 and thyroid stimulating hormone (as measures of thyroid status), and C-reactive protein (as an inflammation marker). Associations of the instrumental variables with miscarriage were investigated using summary association data from women of European ancestry included in our miscarriage GWAS, including 49,996 sporadic miscarriage cases and 174,109 female controls, and 750 recurrent miscarriage cases and 150,215 female controls. We found no significant relationships between either sporadic or recurrent miscarriage and any of the listed chronic health conditions or status-measures. Endometriosis was nominally associated with recurrent miscarriage (Inverse variance weighted odds ratio= 1.03, 95% confidence intervals 1.00-1.07, Pvalue=$2.9x10^{-2}$), but this association did not reach the threshold for multiple testing (Pvalue=$6.25x10^{-3}$). In summary, we find no evidence of a causal link between any of the chronic health conditions tested to date and miscarriage. While data utilised here come from large-scale GWAS including 1000s of individuals, genetic variants significantly associated with each risk factor currently explain small percentages (0.02-5%) of the variance in each trait. Larger GWAS for specific risk factors, and for sporadic and recurrent miscarriage, will be required to clarify some of these potential risk relationships.

PrgmNr 3672 - Development of a discovery pipeline for structural variation contributing to the cause of amyotrophic lateral sclerosis

View session detail

Author Block: **E. McCann**[1], S. Chan Moi Fat[1], L. Henden[1], K. L. Williams[1], G. A. Nicholson[2], D. B. Rowe[1], J. A. Fifita[1], I. P. Blair[1]; [1]Macquarie Univ. Ctr. for MND Res., Sydney, Australia, [2]Northcott NeuroSci. Lab., ANZAC Res. Inst., Sydney, Australia
Disclosure Block: **E. McCann:** None.
Amyotrophic lateral sclerosis (ALS; also known as Lou Gehrigâ□□s disease) is a fatal neurodegenerative disease. Death of the upper and lower motor neurons causes progressive muscle weakness and eventual paralysis, culminating in death from associated respiratory failure. Approximately 10% of ALS cases are familial, while the remainder occur apparently sporadically. Gene mutations are the only known cause of ALS, and with over 30 ALS genes identified, ALS is genetically heterogeneous. Despite heritability estimates of 40-60% for all forms of ALS, only ~40% of familial and ~10% of sporadic ALS can be explained by known ALS gene mutations. As such, almost 90% of ALS cases remain with no known genetic predisposition.

Though almost all ALS gene mutations identified to-date have been small nucleotide level changes, the most common known cause of ALS is a more complex pathogenic expansion of a short tandem repeat within the *C9orf72* gene. This mutation spans thousands of nucleotides, and is therefore comparable in size to large structural variants (SVs). Coupled with the implication of SVs in other neurodegenerative conditions such as Parkinsonâ□□s disease and spinocerebellar ataxia, this suggests that novel SVs are likely to account for a percentage of ALS cases. The study of SVs in ALS has so far been precluded due to various technical hindrances but is now feasible with the advent of sophisticated tools and technologies.

Using whole genome sequencing data from 650 ALS cases, we have developed a discovery pipeline for ALS relevant SVs. This involves the application of several bioinformatics tools including Lumpy, Manta, MetaSV, Duphold, ReciprocalOverlapAnnotator, AnnotSV and Samplot to generate a high confidence SV call set in each individual. This is followed by customised R filtering strategies to remove population-based SVs and perform cohort specific analyses. We have successfully applied this pipeline to a small ALS family to identify just seven candidate SVs from an initial >7000, one of which is potentially acting as the cause of ALS within this family, following the exclusion of all SNP/indel variants. Using these candidate SVs, we are currently optimising our wet-lab SV validation strategy employing a combination of traditional PCR, repeat-primed PCR, TaqMan assays and Bionano Genomics optical genome mapping.

The identification of novel SVs in ALS will not only implicate novel ALS genes, but also highlight SV as an important contributor to ALS pathogenesis. ALS relevant SVs will be prime targets for much needed therapeutic development, particularly given their effects on gene expression and the recent promise shown by gene silencing therapies in ALS clinical trials.

PrgmNr 3673 - First joint exome and metabolome analysis in dyslexia implicates immune system deficits and dysregulated sensory perception

View session detail

Author Block: **R. Nandakumar**[1], V. Dinu[1], X. Shi[1], H. Gu[1], Y. Kim[2], W. H. Raskind[3], B. Peter[4]; [1]Arizona State Univ., Tempe, AZ, [2]Arizona State Univ., tempe, AZ, [3]Univ of Washington, Seattle, WA, [4]Arizona State Univ, Tempe, AZ

Disclosure Block: **R. Nandakumar:** None.

Dyslexia is a common learning disability affecting children's reading and spelling development[1-3]. Dyslexia is thought to be under genetic influence, but despite decades of research, the mechanisms of genetic disruptions are not yet well understood. Similarly unclear is the role of associated health impairments, such as immunodeficiency[4,5] and sensory dysregulation (neural noise hypothesis)[6] Findings are mixed, with some studies showing evidence for immune system disorders [4,7] while others do not [5]. DNA is not the only biological agent that may influence dyslexia. Metabolomics, which is a rapidly evolving field of measuring endogenous metabolites in a cell or body fluid, may contribute new insights into the biology of dyslexia [8]. A few recent studies have examined neurometabolite concentrations in children[9], adults[10], and both[11] with dyslexia. They found a correlation between a wide range of reading skills and neurometabolites concentration (e.g., choline) in medial occipital and left temporoparietal cortices. Here, we report the first experimental joint exome and metabolome study in adults and children with dyslexia and some of their family members. Exomes of 38 individuals (29 with dyslexia) and metabolomes of 26 individuals (17 with dyslexia) were available; both data types of were available for 21 of these individuals. Two metabolites, pyridoxine and kynurenic acid, were found in significantly lower concentrations ($p = .0076, .0258$, respectively) in the individuals with dyslexia, compared to the controls. Pyridoxine plays a role in Vitamin B6 metabolism, which is implicated in autism spectrum disorder (ASD). Kynurenic acid has been implicated in sensory signal/noise filtering in schizophrenia and ASD. A gene ontology analysis based on the output of an allelic association analysis showed significant (p

PrgmNr 3674 - Genome Wide Association Study of QSM brain measures in UK biobank to undercover the genetic basis of brain iron load

View session detail

Author Block: **M. K. Lupton**[1], Z. F. Gerring[2], E. M. Derks[3], S. Macgregor[2], J-S. Ong[2], J. Fripp[4], A. Fazlollahi[4], P. Raniga[4]; [1]Genetic Epidemiology, QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [3]Translational Neurogenomics, QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [4]CSIRO Hlth.and Biosecurity, Australian E-Hlth.Res. Ctr., Brisbane, Australia

Disclosure Block: **M.K. Lupton:** None.

Neuronal iron is likely to be the major source of contrast captured by MRI using quantitative susceptibility mapping (QMS), which quantifies magnetic susceptibility, a fundamental electromagnetic property. Iron accumulation in the brain, as measured by QSM is associated with normal aging and neurodegenerative processes where it has been shown to associate with cognitive decline and implicated in Alzheimerâ□□s disease and Parkinsonâ□□s disease pathogenesis.

Using brain MRI data from the UK Biobank we computed QSM images by combining multi-channel phase and magnitude data from a T2*-weighted sequence with 19.7 ms Echo-time. Anatomical MRI images were automatically segmented to identify corresponding cortical and subcortical regions on QSM.

We carried out a Genome Wide Association Study (GWAS) of neuronal iron measured using QSM in each brain region with a sample size of approximately 35,000 individuals. We identify novel genetic risk variants for iron accumulation in the brain, highlighting key pathways. In addition, we examine genetic relationships of iron accumulation in the brain with peripheral iron measures and neurodegenerative diseases using LD score regression. We investigate causal relationship between iron accumulation and Neurodegenerative diseases using Mendelian Randomization (MR) methods. This is the largest GWAS study to date for neuronal iron measured using QSM. Our findings support the utility of QSM as a proxy for brain iron and increase our understanding of the biological pathways leading to iron accumulation. Previous work has shown a clear association of brain iron accumulation with neurodegenerative disease, but this analysis allows for the investigation of the causal relationship, without confounding from reverse causation.

PrgmNr 3675 - Integrated gene analyses of *de novo* mutations from 46,612 trios with autism and developmental delay

View session detail

Author Block: **T. Wang**[1], C. Kim[2], T. E. Bakken[3], M. A. Gillentine[1], B. Henning[1], Y. Mao[1], C. Gilissen[4], S. Consortium[5], T. J. Nowakowski[6], E. E. Eichler[7]; [1]Dept. of Genome Sci., Univ. of Washington Sch. of Med., Seattle, WA, USA, Seattle, WA, [2]Dept. of Anatomy, Univ. of California, San Francisco (UCSF), San Francisco, CA, USA, San Francisco, CA, [3]Allen Inst. for Brain Sci., Seattle, WA, USA, Seattle, WA, [4]Radboud Univ. Med. Ctr., Nijmegen, Nederland, Netherlands, [5]The Simons Fndn., New York, NY, USA, New York, NY, [6]Dept. of Anatomy, Univ. of California, San Francisco (UCSF), San Francisco, CA, USA; Dept. of Psychiatry and Behavioral Sci., UCSF, San Francisco, CA, USA, San Francisco, CA, [7]Dept. of Genome Sci., Univ. of Washington Sch. of Med., Seattle, WA, USA; Howard Hughes Med. Inst., Univ. of Washington, Seattle, WA, USA, Seattle, WA

Disclosure Block: **T. Wang:** None.

Coding *de novo* mutations (DNMs) have identified hundreds risk genes for neurodevelopmental disorder (NDD); however, most studies have considered autism spectrum disorder (ASD) and developmental delay (DD) separately despite the overwhelming comorbidity and shared DNM genetic etiology. To increase power, we integrated DNMs from 15,560 ASD and 31,052 DD trios and performed comprehensive enrichment analyses using three models. We considered diagnostic overlap, mutation class, sex differences, and functional enrichments of risk genes. In total, we identified 615 NDD candidate risk genes predicted by one or multiple models (FDR 5%, 189 potentially novel), which including 138 genes reaching exome-wide significance supported by all models (*P* GATAD2B and *KIF1A*) are exclusive to have DNM only in DD. In contrast, no ASD-specific genes were identified. Considering the mutation class, we identify 12 genes enriched for likely gene-disruptive DNM compared to 41 genes that incur predominantly missense DNM. By treating males and females separately, we identify 22 genes show sex-specific bias (*P* DDX3X, *MECP2*, *SMC1A*, *WDR45*, and *HDAC8*) also show significant DNM burden in females when compared to males. NDD risk genes grouping into five main protein-protein interaction networks, which tend to associate with distinct cellular lineages based on single-nucleus transcriptomic data. Our results identify new candidate genes as well as functional networks, defining both gene-, mutation class, and sex-specific differences among DNM, which are important for future clinical diagnosis and functional studies.

PrgmNr 3676 - Multivariate genomewide associations on microstructural compartments of human brain highlight the consistent neurodevelopmental molecular pathways from adolescent to late adulthood

View session detail

Author Block: **C. Fan**; Univ California San Diego, La Jolla, CA
Disclosure Block: **C. Fan:** None.
Background: It is important to understand the molecular determinants for microstructures of human brain. Here, we adopt novel imaging processing methods and multivariate GWAS on two large scale imaging genetic datasets to identify key genetic association signals**Study Methods:** UK Biobank participants who have received MRI scans before 2019 as discovery set (n = 23666) and who received MRI scans between 2019 and 2020 were replication set (n = 6396). Furthermore, we used Adolescent Brain Cognitive Development study (ABCD) for further replication (n = 8189). The tissue compartments were inferred based on the mixture modeling [White et al., 2013, 2014]. Three metrics for the tissue compartments were derived. ND as normalized directional signals represents the tissue properties within a tubular structure, such as axonal bodies. N0 as normalized zeroth order signals represents the tissue properties within a cell body, such as neurons. NF as normalized free diffusions represents free water elements outside of the cell body, such as CSF. We applied the combined principal component analyses proposed by Aschard et al., 2014, for discovery. In the replication stage, the loci effects across voxels from the discovery samples were projected onto replication samples. The gene enrichment analyses were performed using MAGMA [de Leeuw et al., 2015] and ShinyGO [Ge et al., 2020]. The anatomical enrichment analyses is by contrasting the observed average effect size per ROI over the null distribution of the average effect size per ROI, sampled from none-significant SNPs (p Results: We found 372, 599, and 471 unique loci for NF, N0, and ND, respectively. Among those discovered loci, 292 (78%), 461 (77%), and 383 (81%) can be replicated in the independent UKBiobank replication set for NF, N0, and ND. Furthermore, 97 of the NF loci (26%), 118 of the N0 loci (20%), and 180 of the ND loci (38%) can be replicated in the ABCD study. The effects of the loci replicated in both datasets are enriched in the cerebellum for NF, corpus callosum for N0, and motor related pathways for ND. Gene-set enrichment analysis shows evident enrichment in neurogenesis and nervous system development across all tissue compartments (Enrichment FDR Conclusion: We found consistent neurodevelopmental molecular pathways from adolescent to late adulthood in microstructures of human brain.

PrgmNr 3677 - Pairwise contributions of 16p11.2 CNV genes to brain-related traits

View session detail

Author Block: **M. Vysotskiy**[1], L. A. Weiss[2]; [1]Dept Psychiatry, Univ California San Francisco, San Francisco, CA, [2]Univ California San Francisco, San Francisco, CA
Disclosure Block: **M. Vysotskiy:** None.
Deletions and duplications of the 16p11.2 copy number variant (CNV) region are associated with a variety of brain-related disorders, including schizophrenia, obesity, autism spectrum disorder, bipolar disorder, and intellectual disability. The region contains approximately 30 coding and 30 noncoding genes, and the relationship between genes at the locus and the associated brain-related traits is not yet understood. We expect deletions and duplications to affect gene expression. To impute gene expression in large (non CNV-carrier) GWAS datasets, and measure association between individual CNV gene expression and trait, we can use PrediXcan. Previously, we have used this methodology to identify individual 16p11.2 genes driving schizophrenia, BMI (measuring obesity), and IQ (measuring intellectual disability), but we could not find single-gene drivers of autism or bipolar disorder. Given that there may be multiple 16p11.2 genes affecting each trait, it may be advantageous to analyze the effects of multiple 16p11.2 genes acting together. To systematically consider pairwise effects of two 16p11.2 genes expressed in the same direction (as would be the case in CNV carriers), we extended the PrediXcan model training and association approach to consider pairs of genes. We tested over 1,500 expression-imputed pairs of 16p11.2 genes for association with autism, BMI, and IQ using publicly available summary statistics across 49 GTEx tissues. We found a contribution of pairs of 16p11.2 genes to BMI larger than that of control genomic regions matched on gene count, length, and coding to noncoding gene ratio. Top associated pairs include those where neither gene had an independent main effect, and those including non-coding genes. Our study provides in-silico evidence that pairwise effects of genes within the 16p11.2 CNV drive one or more brain related traits.

PrgmNr 3678 - Association between complement component 4 variation and childhood brain and behavioral phenotypes: Multiethnic PheWAS in the Adolescent Brain and Cognitive Development (ABCD) Study

View session detail

Author Block: **L. Hernandez**[1], M. Kim[2], G. Hoftman[1], R. Bethlehem[3], W. K. Thompson[4], M. Gandal[5]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [3]Univ. of Cambridge, Cambridge, United Kingdom, [4]Univ. of California, San Diego, La Jolla, CA, [5]UCLA, los angeles, CA

Disclosure Block: **L. Hernandez:** None.

Background: The most significant common variant association for schizophrenia (SZ) reflects increased expression of the complement component 4A (*C4A*) gene. However, it remains unknown whether *C4A* expression affects childhood brain development and is associated with prodromal psychosis. To address these questions, we performed a multiethnic phenome-wide association study in 7,792 youth ages 9-10 to examine how individual variability in genetically predicted *C4A* expression is related to the number and severity of psychosis-like events (PLEs), psychiatric/behavioral traits, and brain structure (volume, cortical thickness, surface area [SA]).

Methods: Data were obtained from the ABCD Study. *C4* haplotypes were imputed using a multiethnic reference panel (Kamitaki, 2020) and predicted *C4A* gene expression was computed using previously published weights (Sekar, 2016). PLEs were measured with the Prodromal Psychosis Scale and the Kiddie Schedule for Affective Disorders and Schizophrenia. Polygenic risk scores (PRS) for SZ were computed using PRS-CS. As previous work suggests that smoking also affects *C4A* expression in the brain (Kim, 2021), we tested how parentâ□□s self-reported smoking may interact with *C4A* to affect childrenâ□□s brain structure. Results were FDR corrected at p**Results**: Predicted *C4A* expression was unrelated to childhood PLEs, psychiatric symptomatology, or global brain size. However, predicted *C4A* expression was associated with reduced SA of the entorhinal cortex (ENT), demonstrating a sex-biased effect such that *C4A* expression had a significantly greater impact on ENT SA in male relative to female youth. Further, smaller ENT SA at 9-10 years predicted, in both males and females, greater number and severity of PLEs 1-year later. SZ PRS was unrelated to ENT SA. The C4A x smoking interaction was significant; youth whose parents smoked more had smaller ENT SA regardless of predicted *C4A* expression levels, while youth whose parents smoked less had smaller ENT SA only in the context of high predicted *C4A* expression.

Discussion: This characterization of the neurodevelopmental effects of *C4A* suggests sex-differences in the effects of *C4A* on regional temporal lobe structure, which is independent of polygenic risk for SZ. Further, we find that the effects of inferred exposure to second-hand smoke on ENT SA mirrors the effects of genetic risk for schizophrenia on the C4 locus. Overall, these data indicate that reduced ENT SA in childhood may serve as a biomarker for schizophrenia risk prior to symptom onset and highlight the importance of large-scale genetic investigations into the predictors of psychiatric disorder in pediatric populations.

PrgmNr 3679 - Polygenic risk and causal inference of psychiatric comorbidity in inflammatory bowel disease

View session detail

Author Block: **P. Hu**[1], Y. Li[2], C. Bernstein[1], W. Xu[2]; [1]Univ. of Manitoba, Winnipeg, MB, Canada, [2]Univ. of Toronto, Toronto, ON, Canada
Disclosure Block: **P. Hu:** None.
Approximately, 40% of patients with inflammatory bowel disease (IBD) experience psychiatric comorbidities (PC). Previous studies demonstrated the polygenetic effect on both IBD and PC. In this study, we evaluated the contribution of the genetic variants to the PC susceptibility among the IBD population, and whether this effect is mediated by the expression levels of *RBPMS*, which is a potential risk factor of PC in IBD patients identified in our previous studies. The polygenic risk score (PRS) was estimated among IBD patients (n=240) from the Manitoba IBD cohort study by using external Genome-Wide Association Studies. The association and prediction performance were examined between the estimated PRS and the PC status in IBD patients. Finally, regression-based models were applied to explore whether the expression levels of the *RBPMS* gene are a mediator between estimated PRS and PC status in IBD. The estimated PRS had a significantly positive association with PC status (for the highest effect: P-value threshold=$5*10^{-3}$, odds ratio=2.3, P-value= $4.1*10^{-7}$). Around 10**%** of the causal effect between the PRS and PC status in IBD was mediated by the expression levels of the *RBPMS* gene. The area under the curve of the PRS-based PC prediction model is around 0.7 at the threshold of $5*10^{-4}$. In conclusion, the PC status in IBD depends on its genetic influences. Around 10% of this genetic influence could be explained by the expression levels of the *RBPMS* gene.

PrgmNr 3680 - Polygenic risk scores for Bipolar Disorder predict age of diagnosis and the number of lifetime psychiatric diagnoses experienced by participants in the Australian Genetics of Bipolar Disorder Study

View session detail

Author Block: **P. A. Lind**[1], I. B. Hickie[2], D. J. Siskind[3], C. M. Olsen[4], D. C. Whiteman[4], N. G. Martin[5], S. E. Medland[5]; [1]Psychiatric Genetics, QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]Brain and Mind Ctr., Univ. of Sydney, Camperdown, Australia, [3]Univ. of Queensland, Brisbane, Australia, [4]Dept. of Population Hlth., QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [5]QIMR Berghofer Med. Res. Inst., Brisbane, Australia

Disclosure Block: **P.A. Lind:** None.

Bipolar Disorder (BD) is a severe mental illness which can result in a high degree of disability. The 2021 Psychiatric Genomics Consortium (PGC) BD genome wide association study (GWAS) identified 64 associated genomic loci (Mullins et al, 2021) and the polygenic risk scores (PRSs) for BD explained ~4.6% of phenotypic variance. Herein we examined whether increased genetic risk (indexed by PRSs derived from the 2021 PGC study) would be associated with earlier age of diagnosis and a higher number of lifetime psychiatric diagnoses in individuals living with BD from a nationally recruited cohort. Genetic and phenotypic data were available for 3,570 unrelated BD cases and 12,591 unrelated controls. Participants with lived experience of BD (68.7% female, aged 44.7Â±13.5, n=3,570) were recruited to the Australian Genetics of Bipolar Disorder study from electronic prescription records and a national media campaign. BD diagnosis was determined using the self-report Mood Disorder Questionnaire (Hirschfield, 2002) and items derived from the DSM-5 criteria for BD. Age of diagnosis (AOD; N=2868) and number of lifetime psychiatric diagnoses (DIAGNOSES; N=3462) were self-reported. Controls were drawn from the Australian QSkin Sun and Health Study (50.9% female, aged 61.3Â±7.9, N=12,591). PRSs were calculated in PLINK using the clumping and thresholding approach. Linear and logistic regressions were performed, controlling for relatedness, ancestry, sex, age and age2 at survey time, and sex*age. The PRS significantly explained 2.8% of the phenotypic variance in BD on the liability scale (P=5.9E-62), 0.07% of AOD (P=8.13E-05), and 0.3% of DIAGNOSES (P=5.9E-04). Compared to individuals in the lowest 10% of the GWAS P-value threshold

PrgmNr 3681 - A GWAS of atopic dermatitis in African American individuals

View session detail

Author Block: **M. Boorgula**[1], M. Daya[1], N. Rafaels[1], M. Campbell[1], S. Chavan[1], C. H. Arehart[1], T. M. Brunetti[1], C. Cox[1], R. A. Mathias[2], K. C. Barnes[1]; [1]Univ Colorado, Aurora, CO, [2]Johns Hopkins Univ, Baltimore, MD

Disclosure Block: **M. Boorgula:** None.

Atopic dermatitis (AD) is a complex chronic skin disease affecting up to 30% of children, often persisting into adulthood. Despite an increased burden of AD in African ancestry populations in the USA, to date, most GWAS of AD were performed in subjects of Asian and European ancestry. The largest GWAS of AD published by the EArly Genetics & Lifecourse Epidemiology (EAGLE) eczema, also included only 1% African ancestry populations. To reduce this paucity of information from GWAS of AD in African ancestry populations, we performed and report the results of a GWAS in 1,113 African American AD cases & controls using Illuminaâ□□s multiethnic global array (MEGA). SNPs that failed on hardy-weinberg, missingness thresholds and that showed significant differences in cases between batches, were excluded. PCA analysis was performed using Genesis PC-AiR and two outliers identified on PC1 (6 standard deviations away from the estimated mean) were excluded. A final data set with 1,111 individuals and 914,377 SNPs was imputed against the TOPMed reference panel and 14,721,903 SNPs with Rsq >= 0.7, MAF>=1% were retained for association analysis. Logistic mixed effects models were used to test for association between allelic dosage and case-control status using the GENESIS R â□¯Bioconductor package. 54 SNPs were suggestive (P-6), from which 13 independent sentinel SNPs were identified. Three SNPs (rs72745472 on chromosome 1, rs12350801 and rs369559451 on chromosome 9) reached genome-wide significance (P-8). SNP rs116000308 (in chromosome 15q22) maps to the *SNX1* gene, which encodes an endosomal protein that regulates the cell-surface expression of epidermal growth factor receptor (EGFR). EGFR plays a crucial role in skin development ,homeostasis and EGFR ligands have been found to be upregulated in chronic inflammatory skin disorders, such as psoriasis, AD, allergic contact dermatitis. All 54 significant SNPs were tested for replication in UKbiobank, EAGLE, BioMe, AD GWAS in Afro-brazilian cohort and imputed data from Colorado Center for Personalized Medicine biobank. Unfortunately, none of the SNPs showed replication in any of these data sets. We identified novel associations in atopic dermatitis individuals from African ancestry in loci on chromosome 15 and chromosome 9 that reached genome level significance. Lack of replication in multiple data sets tested might explain the potential challenge related to power and the definition of the phenotype used in different studies. For future work, we plan to perform integrative omics analyses which will help overcome the power issue and help understand and support novel genetic associations identified from this analysis.

PrgmNr 3683 - A repetitive request: for the genome wide study of short tandem repeats in human phenotypic variation

View session detail

Author Block: **J. Margoliash**[1], S. Saini[2], A. Goren[3], M. Gymrek[2]; [1]Univ. of California, San Diego, CA, [2]Univ California San Diego, La Jolla, CA, [3]UCSD, San Diego, CA
Disclosure Block: **J. Margoliash:** None.
Genome wide association studies (GWAS) have allowed the scientific community to identify the genetic basis of human phenotypes faster than ever before. GWAS findings motivate follow-up mechanistic and medical studies and thus drive our understanding of and ability to treat disease. GWAS traditionally have focused on single nucleotide polymorphisms and short indels (collectively known as SNPs) as they are relatively easy to call from cheap whole genome sequencing (WGS) assays. However, SNPs fail to explain much of the heritability of many complex traits, and detailed analysis of GWAS hits often reveals nearby, more complex, variants to be causal.
Here we produce the first GWAS in a population-scale dataset that studies short tandem repeat (STR) as well as SNPs. STRs are a highly polymorphic class of genetic variants that have been implicated in many diseases and have been estimated to account for 10-15% of *cis* heritability of gene expression. Preliminary evidence suggests noncoding STRs regulate gene expression in a variety of manners - disrupting methylation and repositioning nucleosomes among others. However, until now STRs have largely been excluded from GWAS due to the difficulty of calling them in cheap WGS assays and the lower power inherent in analyzing polymorphic variants, and this has constrained our ability to broadly delineate the role of STRs in human phenotypic expression.
We overcome these limitations by imputing the genotypes from our 2,500 person SNP-STR reference haplotype panel into the SNP callsets of the 500,000 UK Biobank participants. We perform association tests of the imputed STRs against a variety of human phenotypes such as height and white blood cell count. We apply fine-mapping techniques to identify signals previously published by SNP-only GWAS that we predict to be causally driven by a nearby STR, and examine the potential mechanisms by which those STRs act. Lastly, we quantify the additional heritability of each trait explained when also considering STRs.
Our preliminary results suggest that 9% of the genomic signal for height is causally driven by STRs, and that 20% of the genomic regions which are correlated with height contain an STR that has a posterior probability of causality of at least 50%. These data demonstrate that conclusions drawn from GWAS and fine-mapping studies run only on SNPs may frequently be misleading. We hope these data and the analysis pipeline we provide will encourage other researchers to incorporate STRs into their GWAS so as to strengthen the validity of their results, increase their odds of finding medically relevant associations, and provide further insight into the biology of STRs.

PrgmNr 3684 - Enabling biobank-scale GWAS data sharing with efficient compression and indexing

View session detail

Author Block: **K. Schneider**[1], C. R. Gignoux[2], R. Layer[3]; [1]Univ. of Colorado Boulder, Boulder, CO, [2]Univ of Colorado Denver, Anschutz Med. Campus, Aurora, CO, [3]Univ. of Colorado, Boulder, CO
Disclosure Block: **K. Schneider:** None.
A global surge in biobanksâ□□large-scale biomedical databases which match electronic health records to genetic dataâ□□have offered the unprecedented ability to investigate links between phenotype traits and their possible genotypic causes. The ability to use data from these biobanks to uncover important information regarding risk factors and prognosis for a wide range of traits has led to an increase in genome wide association studies (GWAS). Projects such as UK Biobank have done GWAS for hundreds of thousands of individuals across thousands of traits, and we are likely to see similarly sized experiments performed by other biobanks. With current technologies, the summary statistics alone from each GWAS require up to 2GB of storage space, which amounts to over 700TB for the UK Biobank alone. The already large and growing number of GWAS ensures interest in sharing summary statistics data for greater population-wide discovery; but the sizes of these files creates a bottleneck which limits our ability to perform this large-scale sharing and analyses. Current tools like bgzip and tabix are effective for compressing and indexing files of this size, but are general solutions that are not optimized for GWAS data (e.g., summary statistics). For this reason, there is an immediate demand for new methods which allow for especially efficient storage, compute, and sharing ability for GWAS files. We propose a method which achieves more efficient storage with better compression ratio and faster decompression times by use of fast integer-based codecs, which outperform standard codecs (gzip, zlib, or LZ4). Furthermore, we include the addition of summary statistic indexing, which complements traditional genomic position indexing to drastically improve retrieval times of many GWAS queries. This strategy can, for example, quickly identify significant SNPs in a target gene across thousands of traits. Our new file format for compressed GWAS data can be stored, shared, and indexed more efficiently than files compressed with the current standard compression tools, and is freely available at https://github.com/kristen-schneider/gwas-compress.

PrgmNr 3685 - Genetic burden of cardio-metabolic diseases on womenâ□□s health outcomes from a study of multi-ancestry population

View session detail

Author Block: **B. Xiao**[1], A. Lucas[1], T. Drivas[2], M. D. Ritchie[1], S. S. Verma[1]; [1]Univ. of Pennsylvania, Philadelphia, PA, [2]Children's Hosp. of Philadelphia, Philadelphia, PA
Disclosure Block: **B. Xiao:** None.
Cardio-metabolic diseases are generally comorbid with other conditions and are associated with poor health outcomes. However, the investigation of the genetic burden of cardio-metabolic diseases with womenâ□□s health phenotypes such as ovarian cancer and many pregnancy-related complications is highly understudied in large populations. Polygenic risk scores (PRS) can be used to characterize shared genetic effects that could estimate disease risk, and the Penn Medicine Biobank (PMBB) is an electronic health record-based academic biobank consisting of genomic data on many women from multiple ancestries. To analyze the shared genetic effect of PRS from common cardio-metabolic diseases on phenotypes of womenâ□□s health, we calculated PRS for women in PMBB using PRS-CS with an external trans-ancestry LD reference panel and the largest trans-ancestry summary statistics available for seven phenotypes (body mass index (BMI), coronary artery disease, chronic kidney disease, myocardial infarction, type 2 diabetes (T2D), hypertension, and lipid measurements such as high-density lipoprotein (HDL) and triglycerides (TG)). We then tested the association of PRS with clinical lab measurements and case/control phenotypes for all participants and stratified by European and African ancestry. Our analysis identified >30 significant associations reflecting shared biology among common cardio-metabolic phenotypes and diseases such as ovarian cancer, polycystic ovarian syndrome (PCOS), and pregnancy related complications such as postpartum hemorrhage. We replicated known associations such as association of BMI PRS with median overall BMI in women and median BMI during pregnancy. We also identified shared etiology between ovarian cancer and coronary artery disease ($p = 0.0067$) and associations of T2D PRS with PCOS ($p = 0.0064$) and gestational diabetes ($p = 0.0047$) among African ancestry individuals. Additionally, PRS for HDL and TG were significantly associated with excessive fetal growth ($p = 0.0096$) and intrauterine death ($p = 0.0045$) respectively. T2D PRS was also associated with postpartum hemorrhage ($p = 0.002$) when analyzing all participants. Our findings indicate that comorbid womenâ□□s health disorders and cardio-metabolic diseases reflect a dual burden of cardio-metabolic and female-specific genes; the latter may also contribute to pregnancy-related outcomes. Since the strength of association for many phenotypes varied greatly among European and African ancestry groups, our results also highlight that clinical care recognizing racial differences among this dual burden may help improve health outcomes in women living with comorbid disorders.

PrgmNr 3686 - Genetic variants affecting the expression of two Mendelian deafness genes are associated with age-related hearing impairment

View session detail

Author Block: **M. Wilke**[1], B. Drogemoller[1,2,3]; [1]Dept. of Biochemistry and Med. Genetics, Univ. of Manitoba, Winnipeg, MB, Canada, [2]The Childrenâ□□s Hosp. Fndn. of Manitoba, Winnipeg, MB, Canada, [3]CancerCare Manitoba Res. Inst., Winnipeg, MB, Canada
Disclosure Block: **M. Wilke:** None.
Introduction: Age-related hearing impairment (ARHI) is one of the most common sensory impairments in the aging population. ARHI has been shown to be genetically heterogeneous and has genetic mechanisms which are largely unknown. As past genome-wide association studies performed for ARHI have shown that regulatory variants play an important role in ARHI, transcriptome-wide association studies (TWAS) can provide an excellent framework to shed light on the genetic etiology of this complex trait. **Methods:** TWAS were performed on two independent hearing loss cohorts (discovery (n=330,759) and replication (n=52,409)), to determine which hearing loss associated variants have an effect on gene expression levels across various tissues. Both hearing loss genome-wide association summary statistics and expression reference panels from the Genotype-Tissue Expression database were used as input for TWAS analyses performed with S-PrediXcan. Genes that were significantly associated (*P*-6) with hearing loss in the discovery cohort were investigated in the replication cohort. The role of these genes in hearing loss was validated using murine inner ear gene expression databases (gEAR and Shared-Harvard Inner Ear Laboratory Database), gene set overrepresentation analysis (Enrichr) and annotation platforms (FUMA:Gene2Func). **Results:** TWAS analyses in the discovery cohort identified significant associations between hearing loss and thirty-five genes. Two of these genes were replicated in the independent cohort: *TRIOBP* (discovery cohort: $P=4.6x10^{-8}$, z-score=0.67; replication cohort: $P=7.8x10^{-8}$, z-score=1.2) and *ILDR1* (discovery cohort: $P=2x10^{-7}$, z-score=-0.002; replication cohort: $P=4.4x10^{-4}$, z-score=-0.48). Further investigation of these genes revealed that both genes were enriched for abnormal hair cell functions, had high expression in inner and outer ear hair cells, and were associated with hearing loss phenotypes in both mice and humans. **Conclusion:** To the best of our knowledge, this is the first study to perform a TWAS of hearing loss. This TWAS identified two genes, *TRIOBP* and *ILDR1*, that were associated with age-related hearing impairment in two independent hearing loss cohorts. Several lines of evidence have supported the role that these Mendelian deafness genes play in hearing loss. As past research has shown that variants affecting the expression of Mendelian deafness genes play a role in non-Mendelian forms of hearing loss, future research should examine the relevance of these findings to other hearing loss traits.

PrgmNr 3687 - Low-Coverage Whole Genome Sequence Genotyping Pipeline on Heterogeneous Stock Rats

View session detail

Author Block: **D. Chen**, R. Cheng, A. S. Chitre, O. Polesskaya, A. A. Palmer; Univ. of California San Diego, La Jolla, CA
Disclosure Block: **D. Chen:** None.
Both mice and rats are widely used mammalian model organisms; however, for certain phenotypes, rats are preferred over mice. For example, complex behavioral paradigms such as those used to study drug abuse-related phenotypes, are more easily studied in rats. Our laboratory has focused on the use of outbred heterogeneous stock (HS) rats, which were derived in 1984 by intercrossing 8 inbred rat strains, which have now been deeply sequenced. HS rats have been maintained as an outbred population for more than 90 generations and provide both genetic diversity and excellent mapping resolution. In order to perform Genome-Wide Association Studies (GWAS) for drug-abuse related traits, we have developed a cost-effective and high-throughput genotyping pipeline based on low-coverage whole-genome sequencing (LC-WGS). The low-coverage (~0.25x) sequence data are sequenced on Illumina NovaSeq 6000 with Riptide (iGenomX) library preparation, a low-cost whole-genome library preparation kit. Our genotyping pipeline uses BWA for alignment, follows GATKâ□□s best practice for variant calling, and then uses STITCH for genotype imputation. With this LC-WGS genotyping pipeline, we are able to call 8 million SNPs with 99% concordance rate. Moreover, this approach has allowed us to discover tens of thousands of novel SNPs that are not present among the 8 founders. Our laboratory has produced extensive genotype and phenotype data related to drug abuse and other biomedically important traits using these methods.

PrgmNr 3689 - Accelerated epigenetic aging in newborns with Down syndrome

View session detail

Author Block: **K. Xu**[1], S. Li[1], I. S. Muskens[1], N. Elliott[2], S. Myint[1], P. Pandey[1], X. Ma[3], C. Metayer[4], B. A. Mueller[5], K. M. Walsh[6], I. Roberts[2], S. Horvath[7], J. L. Wiemels[1], A. J. de Smith[1]; [1]Ctr. for Genetic Epidemiology, Dept. of Preventive Med., Keck Sch. of Med. of the Univ. of Southern California, Los Angeles, CA, [2]Dept. of Paediatrics and MRC Molecular Haematology Unit, Weatherall Inst. of Molecular Med., Oxford Univ. and BRC Blood Theme, NIHR Oxford BioMed. Ctr., Oxford, United Kingdom, [3]Dept. of Chronic Disease Epidemiology, Yale Sch. of Publ. Hlth., New Haven, CT, [4]Sch. of Publ. Hlth., Univ. of California, Berkeley, Berkeley, CA, [5]Publ. Hlth.Sci. Div., Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [6]Dept. of Neurosurgery, Duke Univ., Durham, NC, [7]Univ California los Angeles, Los Angeles, CA
Disclosure Block: **K. Xu:** None.

Accelerated aging is a hallmark of Down syndrome (DS), with adults experiencing early-onset Alzheimerâ□□s disease and premature aging of skin, hair, and immune and endocrine systems. Accelerated epigenetic aging was found in blood and brain tissue of adults with DS, but when this premature aging begins is unknown. We investigated whether accelerated aging in DS is already detectable in blood at birth.

Dried bloodspots were obtained from 347 newborns with DS and 567 newborns without DS from California or Washington. DNA was isolated, bisulfite-converted, and assayed on Illumina MethylationEPIC DNA methylation (DNAm) arrays. We calculated epigenetic age (DNAmAge) using a published epigenetic clock (391 CpGs) and performed reference-based deconvolution of blood cell proportions using the â□□Identifying Optimal Librariesâ□□ algorithm. *GATA1* was sequenced in a subset of 184 newborns with DS to identify somatic mutations associated with transient abnormal myelopoiesis. We compared DNAmAge between DS and non-DS newborns using linear regression adjusting for chronological age from conception (gestational age plus age at blood sampling), sex, batch, blood cell proportions, and genetic ancestry using EPISTRUCTURE. Age acceleration was calculated as the deviation from expected DNAmAge based on its linear association with chronological age in non-DS newborns. We repeated analyses excluding 61 newborns (60 DS) exceeding mean+1SD for nucleated red blood cell (nRBC) proportions and 30 *GATA1*-positive DS newborns to address potential confounding. We tested for association between *GATA1* mutation variant allele frequency (VAF) and DNAmAge in DS newborns.

Mean chronological age from conception was 269 days in DS and 276 in non-DS newborns. Chronological age was more strongly positively correlated with DNAmAge in non-DS (r=0.17, $P=9.9x10^{-5}$) than DS newborns (r=0.15, P=0.051). Blood cell proportions were significantly associated with DS status and DNAmAge, including strong correlation between nRBCs and DNAmAge (r=0.30, $P=5.8x10^{-19}$). Adjusting for cell proportions, DS was significantly associated with increased DNAmAge (beta=0.2641, $P=8.43x10^{-19}$), with an age acceleration of 245 days. This association remained after excluding high nRBC newborns and *GATA1*-positive DS newborns (beta=0.1865, $P=1.07x10^{-17}$, age acceleration=179 days). Among newborns with DS, *GATA1* mutations were associated with increased DNAmAge ($P=6.65x10^{-12}$), with age acceleration of 110 days per 10% increase in VAF.

Our results support that accelerated aging in blood in DS begins prenatally, with implications for the pathophysiology of immunosenescence and other aging-related traits in DS.

PrgmNr 3690 - Characterizing molecular mechanisms of the oncogene transcription factor FOXM1 in lung and breast cancer to suppress cell proliferation

View session detail

Author Block: **S. Yang**, S. Schreiner, L. Zhu, S. K. Rhie; Dept. of Biochemistry and Molecular Med., Keck Sch. of Med. of Univ. of Southern California, Los Angeles, CA, United States, Los Angeles, CA
Disclosure Block: **S. Yang:** None.
Transcription factors are proteins that bind to DNA to control the rate of transcription. Transcription factors bind to regulatory elements such as promoters and enhancers to control expression of numerous genes. Changes in transcription factor activity could lead to either the inactivation of tumor-suppressor genes, or the activation of oncogenes in cancer. Therefore, characterizing molecular mechanisms of transcription factors dysregulated in cancer and developing specific inhibitors of those transcription factors can provide potential therapeutic candidates.
Using an in-house developed bioinformatic tool called Tracing Enhancer Networks using Epigenetic Traits (TENET), which identifies key transcription factors linked to cell-type specific enhancers with multi-omic datasets generated in tumor tissues, we have identified that the transcription factor FOXM1 is involved in oncogenic networks of tumor-specific enhancers for breast and lung cancer. Moreover, we found that more enhancers linked to FOXM1 are activated in the basal subtype of breast cancer as well as the lung cancer subtype that is associated with worse prognosis.
In this study, we have investigated molecular mechanisms of FOXM1 in both the A549 lung cancer cell line and MDAMB231 basal breast cancer cell line. To identify FOXM1 binding sites, chromatin immunoprecipitation coupled with sequencing (ChIP-seq) experiments were performed. Moreover, we have performed siRNA knockdown followed by RNA-seq experiments to discover genes regulated by FOXM1. We have integrated ChIP-seq data with RNA-seq data to identify target genes of FOXM1 specifically in lung and breast cancer, and identified potential genes directly regulated by FOXM1 in each cancer type. Furthermore, to suppress the activity of dysregulated FOXM1 and to better understand the molecular mechanisms of FOXM1 inhibition, we have treated cancer cells with FOXM1 inhibitors and then performed western blot, RNA-seq, and functional assays that measure cell proliferation rates. We have characterized the effect of FOXM1 inhibitors on the protein expression of FOXM1 as well as the expression of genes that are regulated by FOXM1.
In conclusion, we found that abnormally high expression of FOXM1 alters cancer-type specific enhancer networks in lung and breast cancer, and we are investigating ways to suppress FOXM1 activities in a cancer-type specific manner. Our findings will provide useful information for future development of improved therapeutics as well as diagnostics.

PrgmNr 3691 - Comparative analysis of bulk RNA versus single-cell RNA sequencing to profile gene expression in cellular subgroups within peripheral blood mononuclear cells

View session detail

Author Block: **X. Liu**[1], Z-X. Liu[2], T. R. Morgan[3], T. M. Norden-Krichmar[1]; [1]Univ. of California, Irvine, Irvine, CA, [2]Keck Graduate Inst., Claremont, CA, [3]VA Long Beach Hlth.care System, Long Beach, CA
Disclosure Block: **X. Liu:** None.
Background: The identification of differentially expressed genes between viral- and alcohol-associated hepatitis is critical to understanding the biological mechanisms underlying these two diseases. Recent work from our group has demonstrated that alcohol-associated hepatitis (AH) and chronic hepatitis C viral infection (HCV) can be distinguished by bulk RNA-seq from peripheral blood mononuclear cells (PBMCs). The goal of the current study was to compare expression profiles from single-cell (scRNA) and bulk RNA-seq of PBMCs from the same participants, to determine if there were different cell types driving the gene expression in the PBMCs between these disease groups. **Methods**: Biospecimens from an IRB-approved study by the Southern California Alcoholic Hepatitis Consortium (SCAHC) were collected at baseline with consent by the patients from AH (n=1), HCV (n=1), and healthy control (n=1). ScRNA-seq was performed using the 10x Genomics Chromium Single Cell 3â technology and Illumina sequencing. Sequence data was mapped to the human genome (hg19) and integrated using cellranger 3.1.0 version. We then performed downstream analysis using Seurat (version 4.0.1). Bulk-RNA data was aligned and mapped to the human genome (hg19) using Bowtie2 release 2.3.4.1 and Tophat2 release 2.1.1. Differential expression was then performed using cuffquant and upper quartile normalization by cuffdiff in the Tuxedo suite 2.2.1. Differential expressed genes (DEGs) between the samples were filtered to retain only those with abs(log_2(fold change)) â¥1, FPKM â¥ 5, and significance at FDR-adjusted p-value â¤0.01. **Results**: We compared DEGs across conditions within each of the 9 cell types in the scRNA-seq against the unique significant DEGs between the disease conditions from bulk RNA-seq. Significant DEGs were identified in both bulk and scRNA sequencing. Using scRNA-seq, we were able to detect genes expressed in certain subtypes of cells. For example, interferon-induced gene *IFITM3* was highly expressed in HCV bulk RNA, and was specifically expressed in monocytes and natural killer (NK) cells. The *RETN* gene, which is related to the innate immune system, is highly expressed in AH bulk RNA and was also found to be highly expressed in monocytes. **Conclusion**: This exploratory study assessed the performance of scRNA-seq against bulk RNA using samples of PBMCs from the same participants, so that we could directly compare these two sequencing platforms. Our results provided evidence of transcriptional heterogeneity of cellular subtypes in viral- and alcohol-associated hepatitis.

PrgmNr 3692 - Genome-wide effects of sex chromosome aneuploidy on gene expression and DNA methylation in isogenic neural precursor lines

View session detail

Author Block: **J. Berletch**[1], G. Filippova[1], A. SkakkebÃ¦k[2], W. Ma[3], C. Groneck[1], H. Fang[1], X. Deng[1], C. Gravholt[2], C. M. Disteche[1,4]; [1]Dept. of Lab. Med. and Pathology, Univ. of Washington, Seattle, WA, [2]Aarhus Univ. Hosp., Aarhus, Denmark, [3]Dept. of Statistics, Univ. of California, Riverside, Riverside, CA, [4]Dept. of Med., Univ. of Washington, Seattle, WA
Disclosure Block: **J. Berletch:** None.

Sex chromosome aneuploidy leads to phenotypic abnormalities that often affect brain structure and function. These malfunctions are due to a combination of sex-linked gene imbalance and hormonal dysregulation. For example, in Klinefelter syndrome (XXY), hypogonadism during adolescence and adulthood only partly explain phenotypic abnormalities since many of the co-morbidities are already present at an earlier age, suggesting an important contribution of abnormal gene dosage. Previous studies of sex chromosome aneuploidy done on a limited number of cell types (mainly blood) have detected changes in gene expression and epigenetic features throughout the genome, suggesting widespread effects of aneuploidy. To investigate gene expression and epigenetic changes associated with aneuploidy specifically in neuronal cells, we derived isogenic pairs of human induced pluripotent stem cells (hiPSCs) with a different number of sex chromosomes on the same genetic background. We derived XXY/XY and XXX/X pairs either by reprogramming blood cells from mosaic individuals (XXY/XY and XXX/X) or by removal of the inactive X chromosome in X aneuploid hiPSCs using a selectable marker (thymidine kinase) inserted in the *XIST* gene and selection in ganciclovir. These new isogenic hiPSC lines uniquely position us to address the effects of sex chromosome dosage (SCD), while minimizing variability between individuals as well as environmental or hormonal confounders. We differentiated the isogenic pairs into neural precursor cells (NPCs) to assess epigenetic and transcriptomic consequences of SCD (XXY, XXX, and X). Increased dosage of the sex chromosomes resulted in changes in expression of sex-linked genes, as well as in autosomal genes independent of the sex of the individual, as well as distinct DNA methylation signatures dependent on SCD. Changes were noticeably less prominent between isogenic samples than between independent samples, resulting in identification of critical genes. Together, these data point to trans-acting effects of SCD on autosomal gene regulation in neuronal cells.

PrgmNr 3693 - Interrogation of allergy associated regulatory variants for *GATA3* in human T cells

View session detail

Author Block: **H. Chen**[1], P. C. Fiaux[2], A. Chen[1], G. McVicker[1]; [1]Salk Inst. for Biological Studies, La Jolla, CA, [2]San Diego, CA
Disclosure Block: **H. Chen:** None.
Recent studies suggest allergic diseases may share a common genetic etiology. To characterize the potential common genetic etiology, we analyzed the 38 significant loci from cross-trait analysis on allergic diseases from UK Biobank. GO analysis on all the genes in the 38 loci showed the genes are most enriched in T helper cell differentiation. T helper 2 (Th2) cells are one of the T helper cells that induce allergic inflammation. The differentiation of Th2 cells is regulated by the gene *GATA3*. Carefully examining all 38 loci, we found 2 loci located 1 and 0.6 Mb downstream of *GATA3*, named Risk Region 1 and 2 (RR1 and RR2) respectively. Fine-mapping of the two loci identified the candidate SNPs for the causal variants. Notably, the risk alleles of the causal variants in RR1 have high allele frequency of 0.94. The protective alleles occur more recently and arise after human-chimpanzee split. With the lack of eQTL studies of activated Th2 cells, we inspected all available eQTL studies with relevant cell types. We discovered those variants are associated with changes in *GATA3* expression in naÃ¯ve CD4 T cells, although with much weaker evidence. To uncover all regulatory sequences within 2 Mb region around *GATA3*, we conducted a CRISPR/Cas9 tiling deletion screen in Jurkat T cells. The results analyzed by RELICS discover 36 functional sequences (FSs), most of which are enriched in the *GATA3* TAD. One of the FSs, FS28, marked by strong H3K27ac was found to be a conserved strong enhancer for *GATA3* in mouse T cells. Surprisingly, while some FSs overlap with canonical enhancer signatures, the other FSs are devoid of those signatures in Jurkat cells. FS10, located in RR1, overlaps with the potential causal variants and is one of the unmarked FSs. We validated the screen results by introducing Cas9 RNPs with dual sgRNAs to delete FS28 or FS10 and observe *GATA3* expression. To further characterize the role of FS28 and FS10 for *GATA3* in primary Th2 cells, we intersected the H3K27ac ChIP-seq datasets with the 36 FSs during different T helper cell differentiation. We found both FS28 and FS10 are marked by strong H3K27ac specifically in Th2 cell differentiation. To test whether the loss of FS28 or FS10 affect *GATA3* expression in Th2 cells, we deleted FS28 or FS10 and found decreased *GATA3* expression. Importantly, the deletions of FS28 or FS10 may potentially affect Th2 cell differentiation and maintenance, ultimately leading to loss of Th2 cells. Our results uncover two strong Th2 specific enhancers for *GATA3* and provide Th2 cell specific therapeutic targets. Also, the variants in the enhancers could contribute to the common genetic etiology by altering the activities of the enhancers.

PrgmNr 3694 - Novel insight into the etiology of ischemic stroke gained by integrating human transcriptome wide association study with rodent expression data

View session detail

Author Block: **J. Jung**[1,2], Z. Lu[3], N. Mancuso[2,4]; [1]Dept. of Clinical Pharmacy, Sch. of Pharmacy, Univ. of Southern California, Los Angeles, CA, [2]Ctr. for Genetic Epidemiology, Dept. of Preventive Med., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [3]Biostatistics Div., Dept. of Preventive Med., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [4]Norris Comprehensive Cancer Ctr., Univ. of Southern California, Los Angeles, CA
Disclosure Block: **J. Jung:** None.
Stroke, characterized by sudden neurological deficits, is the second leading cause of death worldwide. The main cause of stroke is ischemic stroke (IS) due to cerebral infarction. Although genome-wide association studies (GWAS) have successfully identified many genomic regions associated with IS, the genes underlying risk and their regulatory mechanisms remain elusive. In contrast, the transcriptome-wide association study (TWAS) approach has allowed for more interpretable biologically relevant results by integrating GWAS associations and functional genomic data. Previous work has identified putative susceptibility genes underlying IS risk through TWAS, however transcriptomic data was limited to whole blood and adipose tissue, which may miss possible disease mechanisms in other disease-relevant tissues.
Here, we sought to identify susceptibility genes for IS risk by integrating large-scale IS GWAS summary statistics (N=440,328) together with expression quantitative trait loci data in 48 tissues from GTEx v7 (median N=166) and CommonMind (N=452), and protein quantitative trait loci data in plasma from the INTERVAL study (N=3,301). We performed TWAS and identified 11 genes in 5 genomic regions (PNext, we performed a proteome-wide association study (PWAS) and identified 3 genes in 3 independent genomic regions (PTo validate our findings, we constructed co-expression networks using a cortex gene expression from IS patients (N=12) together with cortex expression datasets obtained from mouse (N=8) and rat IS models (N=8). We identified a co-expressed gene set consistently expressed in mouse, rat, and human brain cortex linked with the chemical synaptic transmission which is significantly enriched (FDR

PrgmNr 3695 - Prediction of genetically regulated expression of asthma target tissues for African-ancestry populations

View session detail

Author Block: **R. K. Johnson**[1], E. Esquinca[2], C. H. Arehart[3], M. Boorgula[4], M. Campbell[1], S. Chavan[5], N. M. Rafaels[6], C. Cox[7], A. Greenidge[8], P. Maul[8], T. Maul[8], D. Walcott[8], T. M. Brunetti[5], I. Ruczinski[9], K. Kammers[10], H. Watson[8], R. Landis[8], M. Taub[11], M. Daya[4], R. Mathias[9], K. Barnes[12]; [1]Div. of BioMed. Informatics and Personalized Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Colorado Sch. of Publ. Hlth., Dept. of Biostatistics and Informatics, Aurora, CO, [3]Aurora, CO, [4]Univ Colorado, Aurora, CO, [5]Univ Colorado, Denver, CO, [6]Univ. of Colorado Sch. of Med., Aurora, CO, [7]13001 E. 17th Avenue, Aurora, CO 80045, Aurora, CO, [8]Univ. of the West Indies, Queen Elizabeth Hosp., Bridgetown, St. Michael, Barbados, [9]Johns Hopkins Univ, Baltimore, MD, [10]Johns Hopkins Univ Sch. of Med., Baltimore, MD, [11]Johns Hopkins, Baltimore, MD, [12]Univ Colorado Denver, Aurora, CO

Disclosure Block: **R.K. Johnson:** None.

Integration of genetics with transcriptomics has improved our ability to identify genes associated with phenotypes. CD4+ T cells play a central role in modulating allergic disease and asthma. However, this tissue is not well-represented in public repositories such as GTEx, particularly for populations of African ancestry that are disproportionately affected by allergic disease and may have distinct genetic risk factors. From 260 subjects of Afro-Caribbean ancestry participating in the Barbados Asthma Genetics Study (BAGS), we quantified gene expression using RNA-seq from isolated, unstimulated CD4+ T cells. DNA was extracted for genotyping on Illuminaâ□□s Multi Ethnic Global Array and imputed to the TOPMed Freeze5 reference panel. Reproducible workflows/tools implementing the Predixcan family of tools were created on the NHLBI BioData Catalyst Ecosystem powered by Seven Bridges and used to build prediction models for gene expression from genotyping data using nested cross-validated elastic net linear models. We limited potential predictors to cis-acting SNPs within 1Mb up or downstream of the gene location for each gene quantified. Models were fit on gene expression residuals after adjustment for sex, asthma status, genetic PC1, and 45 PEER factors. We achieved significant prediction for 4,072 of 16,692 genes tested, defined as R2>0.01 for predicted versus observed gene expression during nested cross-validation and p

PrgmNr 3696 - Regulatory landscape of CD4+ T cells with implications for asthma

View session detail

Author Block: **C. H. Arehart**[1], M. Taub[2], M. P. Boorgula[1], M. Campbell[1], S. Chavan[1], M. Daya[1], N. Rafaels[1], C. Cox[3], A. Greenidge[4], P. Maul[4], T. Maul[4], D. Walcott[4], T. Brunetti[1], R. K. Johnson[5], I. Ruczinski[6], K. Kammers[7], H. Watson[4], R. C. Landis[4], K. Barnes[8], R. Mathias[6]; [1]Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Johns Hopkins, Baltimore, MD, [3]13001 E. 17th Avenue, Aurora, CO 80045, Aurora, CO, [4]The Univ. of the West Indies, Queen Elizabeth Hosp., Bridgetown, Barbados, [5]Aurora, CO, [6]Johns Hopkins Univ, Baltimore, MD, [7]Johns Hopkins Univ Sch. of Med., Baltimore, MD, [8]Univ Colorado Denver, Aurora, CO

Disclosure Block: **C.H. Arehart:** None.

Asthmatics of African ancestry tend to have more severe asthma than those of European ancestry, but relatively few studies have focused on this underrepresented minority group. In the U.S., childhood asthma prevalence is almost twice as high in African Americans (14.6%) compared to European Americans (8.2%) and in Barbados, prevalence is even higher (~20%) due to the elevated exposure to relevant antigens (house dust mite, domestic endotoxin). The Barbados Asthma Genetics Study (BAGS) provides an opportunity to study the complex genetics underlying asthma susceptibility among individuals of Afro-Caribbean ancestry. Since manifestations of asthma depend on T cell activities, particularly the CD4+ Th2 subset of T cells, we test if genetic variants regulate gene transcription in a cell-specific manner by characterizing the transcriptome of CD4+ T cells from atopic asthmatics and non-asthmatics using RNA-Seq.

We performed expression quantitative trait locus (eQTL) analysis on 190 samples (117 asthmatic cases, 73 controls) genotyped with the Multi Ethnic Global array (MEGA) imputed to the TOPMed Freeze5 reference panel and RNA-Seq on CD4+ T cells. Post quality control and alignment we used MatrixEQTL and eigenMT to select the most significant variant-per-gene. We discovered 3,205 eQTLs, 657 of which were validated in an independent sample drawn from the same population (N = 70; 48 cases, 22 controls) also with RNA-Seq on CD4+ T cells. Comparing the 657 replicated CD4+ T cell eQTLs to 49 Genotype-Tissue Expression (GTEx) v8 tissue types, 66 eQTLs were untested in any GTEx tissue, 60 were tested for and novel to our dataset, 531 showed concordant significance with at least one GTEx tissue, and the overall regulatory landscape was most similar to whole blood. Among known genome wide association study (GWAS) & candidate identified asthma genes, we found eQTLs for *HLAâ□□DQA1*, *HLAâ□□DQB1*, *HLAâ□□DRB1*, *HLAâ□□DPB1*, and *ORMDL3*. Gene ontology analysis for the 657 eQTLs confirmed this cell typeâ□□s relevance to asthma through enrichment for antigen processing/presentation and adaptive immune response biological processes. Analysis of differential expression by asthma status highlighted 338 genes, 6 of which overlapped with replicated eQTLs (*AC015712.6*, *PLEKHH2*, *AC067930.4*, *RBM11*, *TXNRD3*, and *RPL23AP7*).

The present study details the regulatory landscape for CD4+ T cells among African ancestry individuals within the context of asthma. This set of immune-related eQTLs will form a strong basis for our further analyses in tissue types implicated in the biology of asthma such as nasal epithelium.

PrgmNr 3697 - Single-cell RNA sequencing of lung identifies disease-associated cell type-specific eQTLs

View session detail

Author Block: **H. M. Natri**[1], C. B. Azodi[2,3], A. J. Gutierrez[1], L. E. Bui[1], L. Peter[1], R. P. Kendle[1], D. J. McCarthy[2,3], N. E. Banovich[1]; [1]The Translational Genomics Res. Inst., Phoenix, AZ, [2]Univ. of Melbourne, Parkville, Victoria, Australia, [3]St. Vincentâ□□s Inst. of Med. Res., Fitzroy, Victoria, Australia
Disclosure Block: **H.M. Natri:** None.
Genome-wide association studies and functional genomics approaches have identified regulatory loci that contribute to complex traits and disease. However, the effects of expression quantitative trait loci (eQTL) can be both cell-type and context-specific. Single-cell RNA-sequencing (sc-RNAseq) facilitates the mapping of eQTLs across different cell types, allowing the identification of cell type-specific eQTLs that would go undetected by bulk methods. Here we use sc-RNAseq to examine primary human tissue from the lungs of healthy individuals and those with interstitial lung disease (ILD). ILD is a chronic, progressive lung disease characterized by the scarring of lung tissue through epithelial remodeling and accumulation of extracellular matrix (ECM). Pulmonary Fibrosis (PF) is a clinical phenotype that exhibits the end stage of ILD. In general, progression of fibrosis occurs in a gradient resulting in some regions of extreme remodeling and some regions that appear largely normal. For the most common and severe form of ILD, idiopathic PF (IPF), lung transplantation is the only treatment option.
In the current study, we have used sc-RNAseq to generate expression profiles of 532,488 cells derived from 170 lung tissue samples from 52 healthy and 68 ILD donors, including 39 donors with IPF. To identify cell type-specific expression changes associated with genetic variation, we have used Whole Genome Sequencing (WGS) to genotype 97 individuals with available sc-RNAseq profiles. In an effort to map cis-eQTLs across lung cell types, we have used a pseudo-bulk approach aggregating the expression levels of cells for each of the study individuals. To identify shared and cell type-specific signatures, we have performed stratified analyses as well a joint analysis of all cell types to assess cis-eQTL effect size heterogeneity between cell types. Using this approach, we detect clear cell population, lineage, and cell type-specific signals Further integration with GWAS allows connecting PF risk variants to molecular functions in disease-relevant cell types.
These analyses aid in determining the cell types and states in which PF-associated genetic variants function, and highlight the importance of cellular context in gene regulation in health and disease.

PrgmNr 3698 - Association between the tobacco-specific carcinogen metabolite NNAL and DNA methylation in a multiethnic population

View session detail

Author Block: **B. Huang**[1], Y. M. Patel[1], S. E. Murphy[2], A. M. Binder[3], K. D. Siegmund[4], D. O. Stram[4], S. S. Hecht[2], L. Le Marchand[3], S. L. Park[3]; [1]Keck Sch. of Med. of USC, Los Angeles, CA, [2]Univ. of Minnesota, Minneapolis, MN, [3]Univ. of Hawaii Cancer Ctr., Honolulu, HI, [4]Univ of Southern California, Los Angeles, CA

Disclosure Block: **B. Huang:** None.

Introduction: Prior epigenome-wide association studies (EWAS) have found that smoking is associated with differential DNA methylation at numerous CpG loci across the epigenome. To investigate whether these changes are attributable to specific smoking-related constituents, we performed the first EWAS of the tobacco carcinogen metabolite 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL) in a multiethnic population of smokers. We aimed to explore whether these associations differed by race/ethnicity and whether they were independent of smoking dose as measured by total nicotine equivalents (TNE). **Methods:** This study was conducted in a subcohort of participants from the Multiethnic Cohort who were current smokers at time of biospecimen collection, had urinary smoking biomarker data, and never had lung cancer. Total NNAL (sum of NNAL plus its glucuronides) was measured in urine using a validated liquid chromatography-tandem mass spectrometry assay. DNA methylation profiling was performed on serum using the Illumina MethylationEPIC BeadChip assay. We used linear regression to investigate the association between a standard deviation (SD) increase in the natural log-transformed NNAL concentration and DNA methylation beta value in all participants and by race/ethnicity. Additional models adjusting for TNE were run to evaluate the associations of NNAL independent of smoking dose. **Results:** A total of 1,996 current smokers (average age 63.9, 55% female) were included in the analysis. The mean total NNAL concentration was 1.71 pmol/mL and was highest in African Americans (2.44) and lowest in Japanese Americans (1.22). Total NNAL was associated with 88 differentially methylated CpG sites (pâ□□8), many of which were identified in prior smoking-related EWAS. Each SD (0.84) increase in the natural log-transformed total NNAL was associated with decreased methylation at the top significant CpG loci located on *GPR15* (cg19859270), *AHRR* (cg23576855, cg05575921, cg21161138), *PRSS23* (cg14391737), *RARA* (cg17739917), *F2RL3* (cg03636183), *MGAT3* (cg05086879), and an unmapped region on chromosome 2 (cg21566642, cg01940273) (coefficient estimates ranging from -0.004 to -0.032). There was no statistically significant heterogeneity across race/ethnicity. After adjusting for TNE, estimates were attenuated and only two loci on *AHRR* (cg05575921) and *MGAT3* **(**cg05086879) remained statistically significant. **Conclusions:** NNAL is associated with differential methylation at several previously identified smoking-related CpG loci. Further research is needed to establish a causal relationship between NNAL and decreased methylation at *AHRR* and *MGAT3*.

PrgmNr 3699 - Automated High-throughput IP Workflows using Covaris Adaptive Focused Acoustics (AFA) System for Accelerated Epigenomic Mapping

View session detail

Author Block: **A. Goren**[1], G. Heitmann[1], Y. Cau[1], J. Young[1], T. Xu[1], E. Daviso[2], J. Laugharn[2], H. Khoja[3]; [1]UCSD, San Diego, CA, [2]Covaris, Woburn, MA, [3]Covaris Inc,, Woburn, MA
Disclosure Block: **A. Goren:** None.
Mapping the epigenomic landscape is essential for understanding gene regulatory mechanisms along cell fate changes, disease onset and treatment response. Such epigenomic mapping holds high potential for identification of biomarkers and targets for therapy in general, and epigenetic targets in particular (e.g., DNA binding proteins or DNA methylation). Several key methods for studying of the localization cellular biomolecules use Immunoprecipitation (IP) to capture the specific targets. IP builds on the precipitation of a biomolecule such as a protein, modified protein or a modified DNA fragment, by using an antibody that specifically recognizes the target biomolecule.
Yet, existing next generation sequencing (NGS) approaches - such as chromatin immunoprecipitation (ChIP-seq) and methylated DNA immunoprecipitation (meDIP-seq) - require time-consuming multiple-day workflows. They involve many manual steps and are difficult to optimize and standardize for different input material and across laboratories. These limitations hamper the use of such tools for applications that require high robustness, reproducibility, and automatability such as high throughput screening of small molecules in drug development or biomarkers in diagnostics.
To overcome these challenges, we combined our automated ChIP-seq (Busby *et al.*, 2016) with the Covaris Adaptive Focused Acoustics (AFA) technology to enhance and improve the IP process. Our preliminary results provide supporting evidence that micromixing and fluid streaming generated by AFA can highly accelerate the time required to achieve antibody binding while improving stringency and specificity (minimizing non-specific binding).
We tested the ability of AFA to enhance the binding kinetics of antibody-epitope association, improve signal-to-noise ratios, and decrease total processing time. We evaluated both ChIP-seq for a range of epitopes (e.g., H3K4me3, H3K27me3, H3K27ac and CTCF) and meDIP-seq. Our preliminary results show that AFA highly expedites the IP process (Together, we developed a novel use of AFA for accelerated IP that allows enhancing and simplifying IP-based methods such as ChIP-seq and meDIP-seq. The sample preparation, shearing, and IP can be performed in the same 96 well Covaris AFA-TUBE plate and integrated with automated library preparation thus enabling same day sample-to-sequencer workflows. We expect this key advancement to be highly useful for both research, clinical diagnostic applications and discovery pipelines for epigenetics drugs.

PrgmNr 3700 - Black Representation in Genomics Research (BRGR) Study of eQTLs in venous blood of 23andMe research participants with African ancestry

View session detail

Author Block: **K. Fletez-Brant**[1], K. Kukar[2], E. Bullis[2], C. D. Wong[2], J. O'Connell[2], B. Hicks[2], N. Batada[1], A. Petrakovitz[2], 23andMe Research Team, S. J. Pitts[1], M. Moreno[2], V. Vacic[1]; [1]23andMe Therapeutics, South San Francisco, CA, [2]23andMe, Inc, Sunnyvale, CA
Disclosure Block: **K. Fletez-Brant:** Salary/Employment; 23andMe Therapeutics.
African Americans make up about 14% of the US population, but are disproportionately underrepresented in clinical trials. Genomics studies are similarly biased, with the vast majority of gene expression studies having been conducted in European and, to a lesser extent, East Asian cohorts. This underrepresentation hampers research on conditions with high prevalence in African Americans, restricts interpretation of clinical findings, and limits our understanding of variation in human gene regulation, both in a population-specific manner and in general. These limitations reinforce each other in the interpretation of GWAS in African Americans, where the relative increase in genetic heterogeneity makes the use of European functional data suboptimal. This dearth of functional data in African Americans in turn means that GWAS findings unique to African Americans do not receive appropriate follow-up, thereby further increasing health disparities as compared to Europeans. To help remedy this situation, 23andMe, Inc. initiated the Black Representation in Genomics Research (BRGR) eQTL Study, which surveys gene expression and genetic variation in a large cohort of 23andMe research participants with African ancestry. 23andMe contacted research participants (a) who have >50% genetic African ancestry (as determined by our ancestry inference algorithms), (b) reside in the continental US, and (c) are between 20 and 60 years of age. Approximately 1,000 23andMe research participants consented to participate in the study, with over 40% from the South of the United States. Nearly 75% of enrollees were female, and the median age was 37. Saliva-based DNA samples obtained from these participants were subject to whole genome sequencing (WGS), to an average read depth of 20x. Using a mobile door-to-door phlebotomy service, we collected venous blood samples from approximately 800 research participants, which are undergoing RNA-seq at an average 60M reads/sample. To maximize the utility of this study for the advancement of research on health conditions affecting African Americans and better encourage the development of therapeutics that serve to benefit the Black community, 23andMe will make these de-identified RNA-seq and WGS data publicly available to all qualified researchers via dbGaP. We also plan to host a webinar with study participants to share study outcomes and the impact of their contribution. We hopewith this study to both reduce disparities in research with African Americans and to further our understanding of the genetic contribution to gene regulation in humans broadly.

PrgmNr 3701 - Circular RNAs preferentially localize to synapses in the human brain implying critical functions for synaptic plasticity and may be important components of neurodegeneration

View session detail

Author Block: **S. Smukowski**[1], C. Danyko[1], M. Course[2], K. Gudsnuk[1], N. Postupna[1], C. D. Keene[1], P. N. Valdmanis[3]; [1]Univ. of Washington, Seattle, WA, [2]Seattle, WA, [3]Univ of Washington, Seattle, WA
Disclosure Block: **S. Smukowski:** None.
Circular RNAs (circRNAs) are unique transcript isoforms that may play critical regulatory roles in the human brain which could have implications in the progression of neurodegenerative disorders such as Alzheimerâ□□s Disease (AD). CircRNAs are produced by back-splicing a later exon to a preceding exon creating a closed-loop structure. For example, splicing exon 5 in a gene to exon 3, instead of continuing to exon 6, making a loop consisting of exons 3, 4, and 5. CircRNAs can perform a variety of functions, in addition to being translated, including acting as microRNA â□□spongesâ□□ and sequestering RNA binding proteins. They also persist in the cytoplasm due to resistance to exonucleases. It has been previously demonstrated that circRNAs are differentially expressed in postmortem brains of AD patients which suggests that circRNAs could be involved in important regulatory functions that become perturbed during synaptic and neuron degeneration in AD.
As part of our current investigation into synaptic localized translation, we discovered an enrichment of circRNAs located at the synapse. We were able to identify and quantify synaptic-localized RNA via subcellular fractionation of â□□synaptosomesâ□□ from frontal cortex postmortem brain samples of both AD and control patients using a sucrose gradient ultracentrifugation technique and subsequent total RNA sequencing. Not only did we discover localization differences of linear RNAs, but remarkably, we discovered an overwhelming abundance of circRNAs in synaptosomes. Given that there were also significant quantitative differences of these circRNAs between AD and control, we postulate that circRNAs serve important roles in the regulation of synapses that become disrupted during neurodegenerative disease progression. For example, it could be that the presence of circRNAs at the synapse modulates the presence of other transcripts by interacting with microRNA degradation machinery. Critically, several of these candidate genes already have established associations with AD such as *GSK3B* and *AKT3*. Here, we present our work to uncover the significance of these findings which represents a novel investigation into this class of understudied molecules.

PrgmNr 3702 - Dissecting the evolutionary and functional architecture of human enhancer sequences with multiple ancestral origins

View session detail

Author Block: **S. L. Fong**[1], J. A. Capra[2]; [1]Vanderbilt Univ., Nashville, TN, [2]Univ. of California San Francisco, San Francisco, CA
Disclosure Block: **S.L. Fong:** None.
Motivation: Functional divergence of cis-regulatory enhancer sequences is a major driver of vertebrate speciation. However, the relationship between the evolutionary histories of enhancer sequences and their functional constraint is unclear. To address this question, we traced the evolutionary origins of transcribed human enhancer sequences active across diverse cellular contexts. *Results*: While most human transcribed enhancer DNA can be traced to a single evolutionary origin, we estimate that 40% of enhancers are composed of DNA from multiple ancestral origins (i.e. of multiple ages). These â□□multi-originâ□□ enhancers have evolutionary architectures consisting of older â□□coreâ□□ sequences flanked by younger â□□derivedâ□□ sequences. Within multi-origin enhancers, we find that both the derived and core sequences show evidence of independent biochemical enhancer activity. However, specific transcription factors (TFs) have stable preferences for binding core and derived regions that span sequence origins. Despite the evidence for activity and TF binding in both core and derived sequences, derived regions are under lower purifying selection pressures than adjacent cores. As a result, derived regions tolerate more common genetic variation and are enriched for eQTL associated with gene expression variability in human populations. *Conclusions*: We propose that the integration of younger, derived sequences with conserved core sequences generates regulatory substrates with robust enhancer activity across both core and derived regions and the potential for functional variation enriched in younger derived regions. Our analyses demonstrate that considering enhancer evolutionary architectures can aid interpretation of evolutionary forces acting on enhancer sequences and functional variation across human populations.

PrgmNr 3703 - Enhancer enhancer interaction networks involved in long-distance genome regulation link multiple non-coding variants to function

View session detail

Author Block: **X. Lin**; Stanford Univ., Belmont, CA
Disclosure Block: **X. Lin:** None.
Mammalian genomes often use multiple enhancers spanning an ultra long-distance (>Mb) to modulate genes, yet it is not clear how multiple enhancers may coordinate to achieve this task. In addition, Genome-Wide Association Studies (GWAS) reveal that variants in the non-coding regulatory elements, including enhancers, are estimated to account for >90% of the variants in disease. While individual enhancer variants mostly present small-to-modest clinical risks, a combination of multiple genetic variants among various enhancers can greatly amplify the effects of individual low-penetrance variants in complex diseases and traits, suggesting the importance of an enhancer-enhancer interaction network in controlling disease relevant genes and connecting multiple non-coding variants association to function. Recent progress in genomic mapping approaches and CRISPR perturbation technologies have accelerated the identification of enhancer and their roles in gene regulation. However, it remains unclear why multiple enhancers are needed, as well as a quantitative understanding of the underlying enhancer epistasis network, wherein the activity of an enhancer can functionally interact with others.
we used multiplexed CRISPR-based screening with a pooled library consisting of 87,025 sgRNA combinations to combinatorially probe enhancer-enhancer interactions. We discovered a previously uncharacterized two-layer enhancer-enhancer epistasis network across multiple oncogenes. In this network, we defined a class of synergistic regulatory elements (SREs) which are theorized to maintain both expression level and robustness of genes and provide buffering effects against genome instability over long distances. A suite of quantitative experimental and computational approaches, including chromatin interaction assays (Trac-looping), cellular imaging (multicolor 3-dimensional FISH), and machine learning (elastic-net regularized generalized linear model), unveiled mechanisms associated with this enhancer epistasis. We used this information to create a computational model to predict SREs and provided a strategy to link multiple non-coding variants to reveal their epistasis influence in clinical risk. In the genetic association analysis, we found that the SRE variants could cooperatively impact gene expression and alter clinical risk in cancer and autoimmune disorder. Our work unveiled new mechanisms underlying enhancer-mediated control of gene expression in ultra-long genomic distance, with implications for annotation of enhancer function within cells and interpretation of epistasis contribution of non-coding variants in human disease.

PrgmNr 3704 - Human-specific structural variants alter *cis*-regulation and chromatin structure

View session detail

Author Block: **C. Shew**[1,2], G. Kaya[1,2], M. Y. Dennis[1,2]; [1]Univ. of California, Davis, Davis, CA, [2]Genome Ctr., Univ. of California Davis, Davis, CA
Disclosure Block: **C. Shew:** None.
Segmental duplications (SDs) and other structural variants (SVs) comprise most of the variable base pairs within and between primate genomes. Notably, genes within human-specific SDs (HSDs) have known roles in regulating neuronal proliferation and autoimmune response, and SD-mediated rearrangements are associated with a wide variety of neurodevelopmental disorders. We have previously shown that HSD genes exhibit paralog-specific expression patterns across diverse cell and tissue types in spite of their high sequence similarity (>98.5%), likely as a result of *cis*-regulatory divergence following duplication. However, the regulatory landscape of HSDs and other human-specific SVs has yet to be systematically characterized. We have designed a massively parallel reporter assay (MPRA) to quantify *cis*-regulatory activity of 5701 paralogous sequence variants and 2444 chimpanzee orthologs of candidate regulatory elements within HSDs. The MPRA will be performed in human lymphoblastoid cell lines (LCLs) and SH-SY5Y neuroblastoma cells to identify differentially active sequences in those cell types. We have also designed a complementary capture Hi-C assay targeting 2401 promoters within and adjacent to human- and chimpanzee-specific SVs, which will identify candidate enhancers associated with SV promoters as well as changes to promoter-enhancer connectivity across breakpoints. Preliminary analysis of capture Hi-C in human and chimpanzee LCLs has identified human-specific contacts associated with the insertion of *DUSP22B* on chromosome 16. Together, these assays will identify differences in the connectivity and activity of a comprehensive set of regulatory elements associated with HSDs and other SVs in humans. We hypothesize that changes to the expression of recently duplicated and rearranged genes can be modeled with a more complete picture of their regulatory environment. This work will inform how human-specific SVs alter gene expression, with potential insight into neurological and immunological traits unique to our species.

PrgmNr 3705 - Integrative single cell analysis of cardiogenesis identifies developmental cellular trajectories and prioritizes noncoding mutations in congenital heart disease

View session detail

Author Block: **L. Sundaram**[1], M. Ameen[1], K. Wang[1], T. Quertermous[2], A. Kundaje[1], W. Greenleaf[1]; [1]Stanford Univ., Stanford, CA, [2]Stanford Cardiology, Stanford, CA
Disclosure Block: **L. Sundaram:** None.
Congenital heart disease (CHD) is the most common birth defect. CHD is caused by genetic dysregulation of heart development, which involves intricate cell state transitions across numerous precursor cell lineages. However, deciphering causal, de-novo, non-coding CHD mutations has been challenging due to a limited understanding of the dynamic regulatory landscape of developmental trajectories in cardiogenesis. Hence, we profiled the single cell chromatin landscape of human fetal hearts from three early development stages to comprehensively map the cis-regulatory landscape of dynamic regulatory elements and their target genes that manifest all major cardiac cell types across 8 major differentiation trajectories. We trained and interpreted base-pair resolved interpretable deep learning models of cell state specific chromatin landscapes, to systematically infer cis-regulatory motif syntax of cooperative TF binding driving cell state transitions. We used the models to predict and interpret the cell-state specific regulatory impact of non-coding mutations in CHD case and control cohorts. We observed a strong enrichment for deleterious *de novo* mutations from CHD in arterial and capillary endothelium cell, identifying for the first time the relevant cell types causing structural changes to the developing fetal heart. Furthermore, we used iPS derived endothelium cells to validate the regulatory impact of several predicted *de novo* non coding causal mutations, highlighting a viable approach in mapping deleterious mutations disrupting cell-type specific regulatory mechanisms even at the early intractable timepoints of fetal development.

PrgmNr 3706 - Investigating conservation of chromatin accessibility signatures during iPSC reprogramming across human and mouse ATAC-seq data

View session detail

Author Block: **C. Thangavelu**[1], T. M. Norden-Krichmar[2]; [1]Univ. of California Irvine, Irvine, CA, [2]Univ of California Irvine, Irvine, CA
Disclosure Block: **C. Thangavelu:** None.
Background: Induced pluripotent stem cells (iPSCs) can be generated from somatic cells, such as fibroblasts, through a process known as iPSC reprogramming. During reprogramming, dynamic changes occur in chromatin accessibility, which are essential for the conversion of fibroblast cells into iPSCs. iPSCs can be used to generate a wide variety of cells with potential for therapeutic purposes. However, efficiently reprogramming cells to a pluripotent state remains a current challenge, particularly in human cells. To further investigate the differences that underlie species-specific differences in reprogramming efficiency and to identify conserved regulatory patterns between species, we conducted a cross-species meta-analysis of mouse and human ATAC-seq data. **Methods**: In this study, a meta-analysis was conducted using ATAC-sequencing data that was generated as mouse and human fibroblasts transitioned to iPSCs. The data was obtained from the publicly available GEO database (mouse: GSE113431, GSE101905, GSE93029; and human: GSE14764). We followed a similar pipeline with the processed ATAC-seq peak data for each species. Replicates were merged using bedtools intersect, and differential accessibility analysis was conducted to filter out peaks associated with housekeeping genes. Peaks from each timepoint were compared across experiments using the bedtools intersect function. HOMER was used to determine what genes and associated pathways were enriched in these regions. Finally, we compared the results obtained for human versus mouse at the nearest equivalent timepoints for evidence of conservation between these species. **Results**: There were noticeable similarities and differences in the chromatin accessibility signatures as the cells transitioned from fibroblasts into iPSCs in mouse versus human data sets. Accessible peaks were located near genes associated with similar developmental and differentiation-related gene ontology terms during the later human and mouse timepoints. However, there were more differences in the levels of motif enrichment at the later timepoints between the species. **Conclusions**: This study demonstrated that there were consistent trends between the species that point to increased areas of accessibility near genes and motifs associated with pluripotency. However, variation in motif enrichment and reprogramming kinetics were also observed. Taken together, these results and methodology may aid in elucidating species-specific differences in iPSC reprogramming.

PrgmNr 3707 - Systematic reconstruction of the cellular trajectories of mammalian embryogenesis

View session detail

Author Block: **C. Qiu**[1], J. Cao[2], T. Li[1], S. Srivatsan[1,3], X. Huang[1,4], D. Calderon[1], W. S. Noble[1,4], C. M. Disteche[5,6], M. Spielmann[7,8], C. B. Moens[9], C. Trapnell[1,10,11], J. Shendure[1,10,11,12]; [1]Dept. of Genome Sci., Univ. of Washington, Seattle, WA, [2]The Rockefeller Univ., New York, NY, [3]Med. Scientist Training Program, Univ. of Washington, Seattle, WA, [4]Paul G. Allen Sch. of Computer Sci. and Engineering, Univ. of Washington, Seattle, WA, [5]Dept. of Pathology, Univ. of Washington, Seattle, WA, [6]Dept. of Med., Univ. of Washington, Seattle, WA, [7]Human Molecular Genomics Group, Max Planck Inst. for Molecular Genetics, Berlin, Germany, [8]Inst. of Human Genetics, Univ. of LÃ¼beck, LÃ¼beck, Germany, [9]Fred Hutchinson Cancer Res Ctr, Seattle, WA, [10]Brotman Baty Inst. for Precision Med., Seattle, WA, [11]Allen Discovery Ctr. for Cell Lineage Tracing, Seattle, WA, [12]Howard Hughes Med. Inst., Seattle, WA
Disclosure Block: **C. Qiu:** None.

Mammalian embryogenesis is characterized by rapid cellular proliferation and diversification. Within a few weeks, a single cell zygote gives rise to millions of cells expressing a panoply of molecular programs, including much of the diversity that will subsequently be present in adult tissues. Although intensively studied, a comprehensive delineation of the major cellular trajectories that comprise mammalian development in vivo remains elusive. Here we set out to integrate several single cell RNA-seq datasets (scRNA-seq) that collectively span mouse gastrulation and organogenesis. We define cell states at each of 19 successive stages spanning E3.5 to E13.5, heuristically connect them with their pseudo-ancestors and pseudo-descendants, and for a subset of stages, deconvolve their approximate spatial distributions. Despite being constructed through automated procedures, the resulting trajectories of mammalian embryogenesis (TOME) are largely consistent with our contemporary understanding of mammalian development. We leverage TOME to nominate transcription factors (TF) and TF motifs as key regulators of each branch point at which a new cell type emerges. Finally, to facilitate comparisons across vertebrates, we apply the same procedures to single cell datasets of zebrafish and frog embryogenesis, and nominate â□□cell type homologsâ□□ based on shared regulators and transcriptional states.

PrgmNr 3708 - Telomere-to-telomere map of a human epigenome

View session detail

Author Block: **A. B. Stergachis**[1,2], D. Dubocanin[1], A. Sedeno Cortes[1], T. Real[1], J. Ranchalis[1], B. Mallory[1]; [1]Univ. of Washington, Seattle, WA, [2]Brotman Baty Inst. for Precision Med., Seattle, WA
Disclosure Block: **A.B. Stergachis:** None.
Chromatin plays a central role in the regulation of the human genome by modulating the occupancy of DNA-binding proteins, and constructing both repressive and permissive environments for gene regulation. However, our understanding of the chromatin architecture for some of the most structurally important and fastest evolving portions of the human genome remains largely unknown. The completion of a telomere-to-telomere human genome for the CHM13 cell line provides an opportunity for exploring the epigenome within these previously poorly resolved regions. To construct a telomere-to-telomere map of a human epigenome, we leveraged long-read single-molecule chromatin fiber-sequencing (Fiber-seq), which permits the near single-nucleotide mapping of chromatin accessibility, nucleosome positioning, and transcription factor occupancy patterns within portions of the genome previously impervious to short-read epigenomic methods. Specifically, Fiber-seq uses non-specific DNA N^6-adenine methyltransferases to precisely stencil the structure of individual chromatin fibers onto their underlying DNA templates, which are read using PacBio single-molecule long-read sequencing. Application of Fiber-seq to CHM13 cells resulted in chromatin architectures for 1.7 million individual intact chromatin fibers with an average length of 16 kb. The high sequencing quality and long-read nature of these chromatin fibers permitted us to uniquely map the chromatin architecture for 97% of the human genome, including 96% of telomeric, 82% of segmentally duplicated, and 70% of centromeric sequence. Single-molecule telomere stencils precisely reveal the transition between nucleosome-bound DNA and the telomere cap, exposing that the terminal several kilobases of each telomere are devoid of nucleosomes, and that the transition from nucleosome-bound DNA is variably positioned across individual fibers. Although centromeres are predominantly comprised of regularly spaced and highly compacted nucleosome arrays, the kinetochore contains a highly unusual chromatin structure consisting of extremely compacted nucleosomes juxtaposed next to large accessible patches of chromatin mirroring those seen at accessible regulatory DNA. Finally, we expose regulatory elements with divergent regulatory activity across paralogous copies of segmental duplications, revealing segmental duplications as a mechanism for the rewiring of the gene regulatory landscape. Overall, this telomere-to-telomere map of a human epigenome provides a foundation for studying the human epigenome in its entirety.

PrgmNr 3709 - The landscape of alternative polyadenylation in single cells of the developing mouse embryo

View session detail

Author Block: **V. Agarwal**[1], S. Lopez-Darwin[2], D. Kelley[1], J. Shendure[2]; [1]Calico Life Sci., South San Francisco, CA, [2]Univ. of Washington, Seattle, WA
Disclosure Block: **V. Agarwal:** Salary/Employment; Calico Life Sciences.
3â□² untranslated regions (3â□² UTRs) post-transcriptionally regulate mRNA stability, localization, and translation rate. While 3â□²-UTR isoforms have been globally quantified in limited cell types using bulk measurements, their differential usage among cell types during mammalian development remains poorly characterized. In this study, we examined a dataset comprising ~2 million cells spanning E9.5-E13.5 of mouse embryonic development to quantify transcriptome-wide changes in alternative polyadenylation (APA). We observe a global lengthening of 3â□² UTRs across embryonic stages in all cell types, although we detect shorter 3â□² UTRs in hematopoietic lineages and longer 3â□² UTRs in neuronal cell types within each stage. An analysis of RNA-binding protein (RBP) dynamics identifies ELAV-like family members, which are concomitantly induced in neuronal lineages and developmental stages experiencing 3â□²-UTR lengthening, as putative regulators of APA. By measuring 3â□²-UTR isoforms in an expansive single cell dataset, our work provides a transcriptome-wide and organism-wide map of the dynamic landscape of alternative polyadenylation during mammalian organogenesis.

PrgmNr 3710 - Universal DNA methylation age across mammalian tissues

View session detail

Author Block: **A. T. Lu**[1], Z. Fei[2], A. Haghani[1], J. Ernst[3,4,5], K. Raj[6], S. Horvath[1,2]; [1]Dept. of Human Genetics, David Geffen Sch. of Med., Univ. of California Los Angeles, Los Angeles, CA, [2]Dept. of Biostatistics, Fielding Sch. of Publ. Hlth., Univ. of California Los Angeles, Los Angeles, CA, [3]Bioinformatics InterDept.al Program, Univ. of California, Los Angeles, CA, [4]Dept. of Biological Chemistry, Univ. of California, Los Angeles, CA, [5]Eli and Edythe Broad Ctr. of Regenerative Med. and Stem Cell Res., Univ. of California, Los Angeles, CA, [6]Radiation Effects Dept., Ctr. for Radiation, Chemical and Environmental Hazards, Publ. Hlth.England, Didcot, United Kingdom
Disclosure Block: **A.T. Lu:** None.
Aging is often perceived as a degenerative process caused by random accrual of cellular damage over time. In spite of this, age can be accurately estimated by epigenetic clocks based on DNA methylation profiles from almost any tissue of the body. Since such pan-tissue epigenetic clocks have been successfully developed for several different species, it is difficult to ignore the likelihood that a defined and shared mechanism underlies the aging process. To address this, we generated data using 10,528 methylation arrays, each profiling up to 36 thousand cytosines in highly-conserved stretches of DNA, from over 57 tissue-types derived from 142 mammalian species. From these, we identified and characterized specific cytosines, whose methylation levels change with age across mammalian species. Genes associated with these cytosines are greatly enriched in polycomb repressive complex 2-binding sites, encode proteins known to participate in mammalian developmental processes and are implicated in age-related diseases. From the methylation profiles, we constructed three universal clocks that are similarly accurate for estimating ages (r>0.96) of any mammalian species and tissue each with a single mathematical formula. Collectively, these new observations support the notion that aging is indeed evolutionarily conserved and coupled to developmental processes across all mammalian species - a notion that was long-debated without the benefit of this new and compelling evidence.

PrgmNr 3711 - Using Taqman assays to analyze eQTL loci arising from GWAS Studies

View session detail

Author Block: **S. Jackson**, H. Veereshlingham, K. Varma; Thermo Fisher Scientific, South San Francisco, CA
Disclosure Block: **S. Jackson:** Salary/Employment; Thermo Fisher Scientific.
Tremendous progress has been made using genome-wide association studies (GWAS) to link genetic variation with phenotypes and pathologies. In spite of these success, it has been estimated that as many as 90% of SNPs identified in GWAS studies map to non-coding regions, complicating the mechanistic interpretation of the results. It is thought that these non-coding SNPs fall into regulatory regions and influence the expression of a gene or genes. The study of expression quantitative trait loci (eQTLs) provides a method for understanding the link between genetic variants and altered gene expression, and could potentially provide new insights connecting GWAS results to molecular mechanisms. To illustrate how eQTLs can be found and verified, we generated transcriptomic information from various tumor samples using Applied Biosystemsâ□¢â□□ Clariomâ□¢ D microarrays. To find putative eQTLs, we compared the SNPs genotypes and the gene expression levels in these samples to the GTEx database. Potential eQTLs in this set of samples were verified using Applied Biosystemsâ□□ Taqmanâ□¢ SNP Genotyping and Taqman Gene Expression assays. We therefore illustrate a workflow where candidate eQTLs can be confirmed and screened in larger cohorts using the more economical qPCR reagents available from Applied Biosystems.

PrgmNr 3712 - A statistical framework for single-molecule transcription factor footprinting

View session detail

Author Block: **D. Dubocanin**[1], D. M. Witten[2,3], A. B. Stergachis[1,4]; [1]Div. of Med. Genetics, Dept. of Med., Univ. of Washington, Seattle, WA, [2]Dept. of Statistics, Univ. of Washington, Seattle, WA, [3]Dept. of Biostatistics, Univ. of Washington, Seattle, WA, [4]Brotman Baty Inst., Seattle, WA
Disclosure Block: **D. Dubocanin:** None.
Transcription Factor (TF) binding and the regulatory networks they compose form the foundations of gene regulation. Currently, our knowledge on TF binding is based on short-read data, which is unable to delineate TF-DNA interactions in hard-to-map genomic regions, and cannot expose combinatorial dynamics of TF binding along individual chromatin fibers. Here we demonstrate an approach for footprinting individual TF binding events along multi-kilobase single-molecule chromatin fibers. We use a non-specific N6-methyladenine-methyltransferase (m6A-MTase) to methylate all adenine residues in DNA that are not protein-bound, and then perform PacBio sequencing. m6A-modified bases are identified using a Gaussian Mixture Model trained on polymerase kinetics obtained during single-molecule sequencing, and single-molecule nucleosome footprints are identified using a Hidden Markov Model. To identify individual single-molecule TF binding events within nucleosome free regions (NFRs), we developed a statistical framework that adjusts for the intrinsic methylation preferences of the m6A-MTase, as well as the local density of m6A-modified bases. We applied this for the de novo discovery of single-molecule TF footprints, as well as for the identification of single-molecule TF binding events at a priori defined TF elements. This method delineates the combinatorial dynamics of TF binding along multi-kilobase single-molecule DNA fibers at unprecedented resolutions.

PrgmNr 3713 - Analyzing the Effects of GWAS Fine-Mapping on PGS Analyses for Type 2 Diabetes

View session detail

Author Block: **P. Nguyen**, R. Gupta, J. Chin; DNAnexus, Mountain View, CA
Disclosure Block: **P. Nguyen:** Salary/Employment; DNAnexus.

A fundamental goal in precision medicine is to accurately quantify a person's disease risk based on their genomic information. This is important because it can have implications in both medical diagnostics and therapeutic development. For years, researchers have conducted genome-wide association studies (GWAS) to identify variations that are associated with diseases and polygenic score (PGS) analyses to summarize the GWAS information into risk for individuals. However, predictive power can be low, especially for complex diseases with a small number of individuals. Explainability can also be challenging, oftentimes because GWAS variations are not causal variations. Moreover, PGS models are not generalizable across ancestries, limiting their contributions to individuals with well-studied ancestries. Here, we analyze the effects of GWAS fine-mapping in terms of mitigating these issues.

For illustration, we focus on quantifying risk for type 2 diabetes (T2D), a disease that affects millions of people in the USA, with 200,000 reported cases per year. We build on top of a scalable framework backed by Apache Spark (which we presented last year) in order to conduct GWAS and PGS analyses at scale for 500,000 participants and 90,000,000 variations from the UK Biobank. We leverage EpiMap, a curated epigenomic dataset consisting of 10,000 epigenomic profiles covering 800 tissues and a number of histone modifications, in order to perform GWAS fine-mapping. We show that GWAS fine-mapping can improve predictive power and generalizability across ancestries. Furthermore, we demonstrate that it can suggest possible pathways that contribute to T2D, which can improve explainability.

All of our analyses are powered by the UK Biobank Research Analysis Platform, a cloud platform for genomics research and translational informatics. Each step is packaged as a workflow or notebook and can be made available upon request. Our analyses are generalized and can be applied to other phenotypic traits at will. We hope our analyses would enable not only more accurate and explainable risk quantification, but also in a way that is generalized and can be useful for more individuals across different ancestries.

PrgmNr 3714 - ChromBPNet: Deep learning models of base-resolution chromatin profiles reveal cis-regulatory syntax and regulatory variation

View session detail

Author Block: **A. Pampari**[1], A. Shcherbina[2], A. Kundaje[3]; [1]Computer Sci., Stanford Univ., Stanford, CA, [2]BioMed. Data Sci., Stanford Univ., Stanford, CA, [3]Stanford Univ., Stanford, CA
Disclosure Block: **A. Pampari:** None.

Chromatin profiling experiments such as DNase-seq, ATAC-seq, and histone ChIP-seq decorate cis-regulatory elements (cREs) with intricate read coverage profiles, whose magnitude, shape and span is regulated by cooperative binding of transcription factors (TFs). Here, we introduce ChromBPNet, the first end-to-end deep learning framework to map DNA sequence to base-resolution chromatin profiles, decipher predictive cis-regulatory sequence syntax of individual cREs and predict the impact of regulatory genetic variants on the strength and shape of different types of chromatin profiles across multiple cell types.

ChromBPNet is an optimized convolutional neural network architecture that models the influence of wide genomic sequence contexts (2-75 kbp) on quantitative base-resolution regulatory profiles from DNase-seq, ATAC-seq and histone ChIP-seq experiments. ChromBPNet trained on five ENCODE canonical cell lines achieved unprecedented predictive performance in held-out chromosomes, while automatically and optimally regressing out assay biases (DNase, Tn5 enzyme bias and input control for ChIP-seq). The models are highly performant over a range of sequencing depths and are able to de-noise and de-sparsify low coverage signal profiles at individual cREs.

We improved interpretation methods for de-novo inference of contribution of individual nucleotides across all putative cREs in the genome, thereby revealing predictive motif instances and their combinatorial interaction effects on base-resolution profiles. We deciphered syntactic sequence heterogeneity of all cREs in each cell-line and benchmarked these predictions against footprinting methods and TF ChIP-seq data. We found remarkable consistency between syntax derived from DNase-seq and ATAC-seq experiments. However, we also found intriguing differences in the influence of motif syntax and their associated cooperative TF complexes on different layers of regulatory activity (TF binding, chromatin accessibility, different types of histone marks) of individual cREs.

Finally, we developed a new variant effect score which predicts the impact of non-coding variants on the strength and shape of base-resolution chromatin profiles, thereby revealing a range of 'blast radii' of variants disrupting different types of TF motifs and syntax. Our models accurately predict quantitative trait loci associated with binding, accessibility and histone marks in lymphoblastoid cell-lines.

Our framework will enable high-resolution annotation of sequence syntax and regulatory variants in 100s of cell types that have been profiled with DNase-seq, ATAC-seq, and histone ChIP-seq experiments.

PrgmNr 3715 - Deep Learning Modeling of Transcription Factor Binding Specificity Using DNA Biophysical Properties

View session detail

Author Block: **B. Alexandrov**[1], M. Bhattarai[1], R. Vangara[1], S. Qian[2], T-P. Chiu[3], K. Rasmussen[1], A. R. Bishop[1], R. Rohs[3], J. Duan[4], X. He[5]; [1]Los Alamos Natl. Lab., Los Alamos, NM, [2]Univ. of Chicago, Chicago, IL, [3]Univ. of Southern California, Los Angeles, CA, [4]Northshore Univ Hlth.system/Univ of Chicago, Evanston, IL, [5]Univ Chicago, Chicago, IL

Disclosure Block: **B. Alexandrov:** None.

Active transcription is initiated and assisted by transcription factors (TF) binding to DNA, a process influenced by various epigenomic mechanisms. Local biophysical properties of DNA, such as local thermodynamic stability, shape, and flexibility, are important for TF-DNA binding. The structural stability of DNA is primarily governed by hydrogen bonds. Because hydrogen bonds are much weaker than covalent bonds, DNA at physiological temperature experiences conformational motions resulting from inherent thermal fluctuations. Such thermal motions of double-stranded DNA spontaneously induce local opening and closing of the double helix, i.e., â□□DNA breathing dynamicsâ□□ or â□□DNA bubblesâ□□ and their propensities are interrelated with DNA local flexibility. The effects of DNA breathing features, such as base-pair flipping propensity and typical bubble length and height on TF binding and gene expression have been shown by multiple studies. DNA shape (i.e., the local 3D structure of the double helix) is also important for TF binding. Many TFs not only recognize DNA sequence but also local DNA shape features, such as DNA bending and minor groove width. Many of the breathing and shape features can be influenced by nucleotides outside the TF motif, which has the potential to shed light on mechanisms of functional noncoding variants associated with disease risk, found in numerous genome-wide association studies. In this work, we integrate biophysical properties of DNA, derived in a high-throughput manner with DNA sequence into the modeling of TF binding specificity in a deep attention DNA machine learning model (DADm). We trained our DADm on TF binding specificity using protein binding microarray data for 203 mammalian TFs. We demonstrate that DADm, based only on DNA sequence, outperforms other state-of-the-art methods. Further, we encode in DADm DNA breathing dynamics and shape together with the sequence. Thus, DADm convolution layer captures the sequence along with DNA biophysical characteristics, the attention layer captures their significant importance and the recurrent layer captures long-term dependencies between the sequence and physicochemical characteristics to learn a new regulatory 'grammar' and improve DADmâ□□s predictions. Our results show that DNA breathing and shape characteristics-augmented models compared favorably to sequence-based models, which we also verified with attention maps of the trained model.

PrgmNr 3716 - DMRscaler: A scale-aware method to identify regions of differential DNA methylation spanning basepair to multi megabase features

View session detail

Author Block: **L. Bondhus**[1], A. Wei[1], V. A. Arboleda[2]; [1]UCLA, Los Angeles, CA, [2]Univ California los Angeles, Los Angeles, CA
Disclosure Block: **L. Bondhus:** None.
Pathogenic mutations in genes that control chromatin function cause epigenetic aberrations resulting in rare genetic syndromes. These chromatin modifiers exhibit extraordinary diversity in the scale of the epigenetic changes they affect, from single basepair modifications by DNMT1 to whole genome structural changes by PRM1/2. Patterns of DNA methylation are related to a diverse set of epigenetic features across this full range of epigenetic scale, making DNA methylation valuable for mapping regions of general epigenetic dysregulation. However, no existing methods make use of these relations to accurately identify the scale of epigenetic changes directly from DNA methylation data. To address this, we developed DMRscaler, a novel method that uses an iterative windowing procedure to capture regions of differential DNA methylation (DMRs) ranging in size from single basepairs to whole chromosomes. We benchmarked DMRscaler against several DMR callers in simulated and natural data comparing XX and XY peripheral blood samples. DMRscaler was the only method that accurately called DMRs ranging in size from 100 bp to 1 Mb (pearson's r = 0.92) and up to 152Mb on the X-chromosome. We then analyzed methylation data from rare-disease cohorts that harbor chromatin modifier gene mutations in NSD1, EZH2, and KAT6A. DMRscaler identified novel DMRs spanning PCDHA, PCDHB and PCDHGB gene clusters across these three groups suggesting these are common mechanisms driving their epigenetic dysregulation. Taken together, our results show DMRscaler is uniquely able to capture the scale of DMR features and identify novel, co-regulated regions that drive epigenetic dysregulation in human disease.

PrgmNr 3717 - Outlier gene detection to improve biological validation and classification performance in RNA sequencing

View session detail

Author Block: **S. Listopad**, T. M. Norden-Krichmar; Univ. of California, Irvine, Irvine, CA
Disclosure Block: **S. Listopad:** None.
Background: RNA sequencing (RNA-seq) serves as a proxy for the level of gene expression in a biological sample. One challenge with interpretation of RNA-seq output, however, involves expression of non-coding genes that were presumed to be removed via poly(A)-selection. It is also common to observe genes with aberrant expression that poorly distinguish between the study conditions, thereby hindering classification performance. The removal of such genes may also improve biological validation of identified gene signatures, since protein coding genes are more extensively annotated than non-coding ones. **Methods:** For this study, we used RNA-seq data obtained from 38 individuals with alcohol-associated hepatitis and 20 healthy controls. The RNA-seq data was aligned to the human genome hg38 using the STAR software and ENSEMBL gene annotation. We generated distribution normalized counts using Cuffnorm for every sample. A gene was labeled as an outlier if it was both non-coding and aberrantly expressed. The non-coding status was determined using ENSEMBL database. A gene was categorized as aberrant if the counts for at least one sample were further than a standard deviation threshold from the mean. We then developed a feature selection and classification pipeline that accepts RNA-seq counts as input, and outputs gene signatures along with corresponding classification performances. We estimated the biological validity of the resultant gene signatures using Enricher. We executed the pipeline with and without outlier gene removal, and then compared the resulting biological validation and classification accuracies. **Results:** Detection and removal of outlier genes had minor effects on classification performance. The classification accuracy varied by Â± 2% depending on the underlying feature selection architecture and choice of the machine learning classifier. The biological validation, however, improved notably with outlier removal. Typically, the number of relevant Enricher hits increased by a factor of 1.4. **Conclusion:** While the outlier removal procedure had negligible effect on classification performance it significantly improved the biological relevance of resulting gene signatures. It is likely that the classification performance did not consistently improve because some aberrant genes are still effective at distinguishing between the conditions. The effect on biological relevance, on other hand, benefited from protein-coding gene bias in existing literature.

PrgmNr 3718 - A Genealogical Estimate of Genetic Relationships to Improve Detection of Population Structure Over Time

View session detail

Author Block: **C. W. Chiang**, C. Fan, N. Mancuso; Univ. of Southern California, Los Angeles, CA
Disclosure Block: **C.W. Chiang:** None.

The application of genetic relationships among individuals, characterized by a genetic relationship matrix (GRM), has far-reaching effects in genetic epidemiology. However, the current standard to calculate the GRM does not take advantage of linkage information and does not reflect the underlying genealogical history of the study sample. Here, we propose a coalescent-informed framework to infer the expected relatedness between pairs of individuals given an ancestral recombination graph (ARG) of the sample. This expected GRM (eGRM) is an unbiased and highly correlated estimate ($r^2 > 0.97$) of the latent pairwise genome-wide relatedness and maintains the mathematical properties of canonical GRMs. Through extensive simulations we show that the eGRM is robust when using genealogies inferred from incomplete genetic data, and can reveal the time-varying nature of population structure in a spatial sample. When applied to genotyping data from a population sample from Northern and Eastern Finland (N=2,644), we found that clustering analysis using the eGRM more accurately delineates population structure than would be possible using the standard GRM, and the temporal pattern of population structure in this sample is consistent with that of a recently diverged and expanded population. Taken together, our proposed estimator drastically shifts the notion of genetic relatedness from a variant-centric to a tree-centric world view, and will be widely applicable to genetic studies in understudied human or ecological samples where whole genome sequencing data or references might not be readily available.

PrgmNr 3719 - Causal variant effect sizes of complex traits differ between populations

View session detail

Author Block: **S. Musharoff**[1,2], R. Patel[1], J. Spence[1], H. Pimentel[3], C. Tcheandjieu[4,2], H. Mostafavi[1], N. Sinnott-Armstrong[1,2], S. L. Clarke[4,2], C. Smith[1], VA Million Veteran Program, P. P. Durda[5], K. D. Taylor[6], R. Tracy[5], Y. Liu[7], C. Johnson[8], S. S. Rich[9], J. I. Rotter[6], F. Aguet[10], K. G. Ardlie[10], S. Gabriel[10], D. Nickerson[8], J. D. Smith[8], P. Tsao[4,2], S. B. Montgomery[1], T. Assimes[4,2], J. Pritchard[1]; [1]Stanford Univ., Stanford, CA, [2]VA Palo Alto Hlth.Care System, Palo Alto, CA, [3]Univ. of California, Los Angeles, Los Angeles, CA, [4]Stanford Univ. Sch. of Med., Stanford, CA, [5]Univ. of Vermont, Burlington, VT, [6]Lundquist Inst., Harbor-UCLA Med Ctr, Torrance, CA, [7]Duke Molecular Physiology Inst., Durham, NC, [8]Univ. of Washington, Seattle, WA, [9]Univ. of Virginia, Charlottesville, VA, [10]Broad Inst. of MIT and Harvard, Cambridge, MA

Disclosure Block: **S. Musharoff:** None.

Despite the growing number of genome-wide association studies for complex traits, it remains unclear whether effect sizes of causal genetic variants differ between populations. Effect sizes of causal variants can differ between populations due to gene-by-gene or gene-by-environment interactions, which have important implications for the study of complex traits and for the use of polygenic scores. However, comparing causal variant effect sizes is challenging: causal variants are hard to identify, and comparisons of their tag SNPsâ effect sizes are confounded by differences in linkage disequilibrium (LD) patterns and allele frequencies between ancestries. Here, we develop a method to assess causal variant effect size differences that does not require identifying the causal variants themselves. Specifically, we leverage the fact that segments of European ancestry shared between European-American and admixed African-American individuals have the same LD patterns and allele frequencies, such that comparisons of tag SNP effect sizes in European ancestry segments are not confounded by these factors. We apply our method to two traits, gene expression of 499 European-Americans and 319 African-Americans in the Multi-Ethnic Study of Atherosclerosis (MESA) and blood lipid levels (statin-adjusted maximum low-density lipoprotein cholesterol, or LDL-C) of ~273K European-Americans and ~72K African-Americans in the Million Veteran Program (MVP). First, we find that global ancestry, local ancestry, and genotype-by-local-ancestry interactions all significantly contribute to the phenotypic variance of both traits, demonstrating the complex association between ancestry and trait architecture. Next, we find that effect sizes of causal variants differ between European-Americans and African-Americans for gene expression and blood lipid levels, likely due to gene-by-gene or gene-by-environment interactions. For gene expression, we find that differences in causal variant effect sizes explain 14% of the decrease in the cross-population genetic correlation. These differences highlight the role of genetic interactions in trait architecture and may contribute to the poor portability of polygenic scores across populations, reinforcing the importance of conducting GWAS on individuals of diverse ancestries and environments.

PrgmNr 3720 - Codon bias analysis of wildtype SARS-CoV-2 and prevalent variants indicates a mutational trajectory independent of human pulmonary tissue and genomic bias

View session detail

Author Block: **S. Brugger**, M. Grose, J. Grose, M. Davis; Brigham Young Univ., Provo, UT
Disclosure Block: **S. Brugger:** None.
Codon bias is a genomic phenomenon in which an organismâ□□s DNA coding regions exhibit a clear usage bias among synonymous codons. Every organism possesses a unique bias, and variation in this bias has even been observed between an organismâ□□s tissue systems. The aim of this study was to determine whether prominent variants of SARS-CoV-2 present evidence of mutational pressure to align their codon biases with those observed both within pulmonary tissue and throughout the human genome. To establish a baseline for analysis, the codon bias of the SARS-CoV-2 reference sequence (WT) was calculated and cross-referenced with the two aforementioned human biases. Despite a common bias for codon AAT (N) between WT and the human genome, no other biases were shared. In order to track the potential development of shared bias, three analyses were performed evaluating a) human genomic and pulmonary bias against prevalent SARS-CoV-2 variants, b) changes in the bias of said variants over time, and c) WT and human biases against the most recently sequenced variants globally. Results of the first analysis revealed modest shifting toward human biases, most notably within B.1.429 and B.1.351 when compared to pulmonary bias. Results of the second analysis showed little shifting from month to month across all tested variants; despite several instances in which samples did demonstrate a shift in bias to align with the human biases, those shifts were not sustained. Variant B.1.429 experienced a significant shift in bias from GTG (V) to GTT (V) from 10/2020 to 12/2020 and again from 02/2021 to 04/2021; similarly, a shift from GAC (D) to GAT (D) was observed over the same time period. Bias for the pre-shift codons was shared between B.1.429 and the two human biases, and the observed shift realigned B.1.429â□□s bias with that of WT. Despite marked preference shifts in several countries and codons, the results of the third analysis did not reveal any long-term changes to codon preference in comparison to WT. Of note, the WT bias for GTT (V) over GTG (V) was generally sustained in this analysis, but that shift was reversed in four samples and a trend toward reversal was observed in an additional five samples. Overall our results indicate that over time, and across several prominent and novel variants, SARS-CoV-2 does not appear to have significantly shifted its codon bias to align with those of the human genome or pulmonary tissue, indicating that such mutations are unnecessary for effective replication and transmission. Research into anomalous codon shifts, such as those observed in valine, may illuminate genomic mechanisms that influence the success or failure of the virus and warrant further investigation.

PrgmNr 3721 - Coevolution is pervasive between unrelated glycosylation pathways and points to potential disease modifiers

View session detail

Author Block: **H. Thorpe**, N. Clark, C. Y. Chow; Univ. of Utah, Salt Lake City, UT
Disclosure Block: **H. Thorpe:** None.
Protein glycosylation is the most common post-translational modification, including N-linked glycosylation, O-linked glycosylation, and GPI anchor biosynthesis. These pathways are separate with nearly no overlapping components. Because of the ubiquity of protein glycosylation, mutations in these pathways lead to a variety of multisystemic metabolic disorders broadly classified as Congenital Disorders of Glycosylation (CDG). CDGs are inherited defects in one of the >150 genes involved in the glycosylation pathways. CDGs can present with a range of symptoms, but the most consistent are seizures, hypotonia, and developmental delays. Glycosylation is ubiquitous throughout the body and affects many proteins making it difficult to identify effective treatments. As with all Mendelian diseases, clinical variability between CDG patients, even with the same mutation, is common and suggests that modifier genes are affecting the phenotypes. I am employing evolutionary approaches to identify modifier genes of CDGs. Evolutionary Rate Covariation (ERC) relies on the premise that proteins that interact physically or genetically or are functionally related coevolve at similar rates. ERC values are calculated using the correlation coefficient of evolutionary rates of gene pairs in a species tree. I pulled pairwise ERC values for 224 genes involved in protein glycosylation calculated from a species tree with evolutionary rates from vertebrates, worms, *Drosophila*, and yeast. As expected, I found enrichment for high ERC values within each glycosylation pathway. Surprisingly, I also identified several components of each pathway that showed high ERC with unrelated glycosylation pathways. For example, MAN2A2, a component of N-glycosylation localized to the Golgi, shows strong ERC with most components of the GPI anchor pathway in the ER. These types of coevolution signatures tell us that there are unappreciated connections between these unrelated pathways. I have also used ERC to look genome wide for coevolution with CDG genes. For example, I pulled ERC values for each GPI anchor synthesis gene across the genome. Gene Ontology analysis of genes that coevolve with GPI anchor synthesis genes showed enrichment of genes involved in RNA modification and mitochondrial gene expression. This enrichment indicates there may be interactions between these processes and GPI anchor synthesis. There was also an enrichment of genes involved in N-linked glycosylation, supporting the connections found previously between GPI anchor synthesis and N-linked glycosylation. These connections could lead to better understanding of the pathways and potential treatments for CDGs.

PrgmNr 3722 - DNA methylation patterns underlying lifespan differences in mammals

View session detail

Author Block: **A. Haghani**, S. Horvath; Univ. of California Los Angeles, Los Angeles, CA
Disclosure Block: **A. Haghani:** None.
The comparative cross-species analysis is a powerful tool to resolve the mysteries of evolution and phenotypic disparities among animals. This study describes the largest multi-species DNAm dataset that was collected by over 100 collaborators from Mammalian Methylation Consortium. This dataset includes over 10,000 DNA methylome data from multiple tissues of different age ranges of over 190 mammalian species. The network analysis of this dataset allowed us to identify co-methylation modules that relate to the individual (age, sex, tissue type) and species characteristics (e.g., phylogenetic order, maximum lifespan, adult weight). The unexpected correlation between DNA methylation and species was sufficiently strong to construct *phyloepigenetic* trees that parallel the phylogenetic tree. The analysis identified the epigenetic marks and the associated genes that relate to the maximum lifespan of mammals. Moreover, we could define novel epigenetic biomarkers of longevity that responded to gold standard anti-aging interventions in mice such as caloric restriction or growth hormone receptor knock outs. Our novel cross-species epigenetic analysis is a rich source of targets for future experimental studies of aging and longevity.

PrgmNr 3723 - Gnomix: local ancestry inference for whole genome sequences and large datasets

View session detail

Author Block: **A. Kumar**[1], H. Hilmarsson[1], R. Rastogi[2], D. Mas Montserrat[1], C. D. Bustamante[3], A. Ioannidis[1]; [1]Stanford Univ., Stanford, CA, [2]Cornell Univ., Ithaca, NY, [3]Stanford Univ. Sch. of Med., Stanford, CA

Disclosure Block: **A. Kumar:** None.

Local ancestry inference is becoming an increasingly important tool as genetic studies become more inclusive. In particular it is now used as a pre-processing step, when running genome wide association studies on diverse cohorts. We introduce Gnomix, a faster, more accurate, high resolution local ancestry inference algorithm designed for the challenges of modern biobank-scale datasets and whole genome sequences. Gnomix has a modular approach allowing the user to choose from a host of base learners and smoothers to optimize the tradeoff of performance vs. accuracy for any particular application. Because the Gnomix model, once trained, can be saved and re-deployed, Gnomix allows for training once with future inference able to be performed rapidly without the need for re-training. Since the trained models are portable, Gnomix allows collaborators to share compact inference algorithms, pretrained on large datasets comprised of samples that may themselves not be easy to share either due to the difficulties of moving data or due to restrictions on the transfer or release of the raw individual sequences in the datasets. This permits highly accurate models to be trained on large volumes of training data held in one location and then applied across multiple independent studies, saving time while increasing accuracy and reproducibility. We demonstrate Gnomix on worldwide datasets of whole genome human sequences, exhibiting its performance metrics on human sequences from the NHGRI Genome Sequencing Program (GSP), as well as on other species. In all cases we show robust improvements of both accuracy and speed (more than an order of magnitude), when compared to a panel consisting of currently popular methods.

PrgmNr 3724 - Human-specific evolution at multiple scales and dimensions

View session detail

Author Block: **K. Keough**, K. S. Pollard; Gladstone Inst.s, San Francisco, CA
Disclosure Block: **K. Keough:** None.
Human Accelerated Regions (HARs), conserved genomic loci that evolved at an accelerated rate in the human lineage, are of intense interest based on their potential to underlie human-specific traits. We optimized an open-source Nextflow pipeline to identify lineage-specific accelerated regions in alignments of hundreds of species and used it to identify an updated set of HARs and chimpanzee accelerated regions. We observed a striking enrichment of human lineage-specific structural variants in topological associating domains containing HARs. Combining machine-learning predictions with chromatin capture experiments in human and chimpanzee neural progenitor cells, we discovered human-specific changes to three-dimensional genome organization around HARs, potentially rewiring regulatory interactions with neurodevelopmental genes. Thus, comparative genomics together with 3D genome organization data revealed enhancer hijacking as a mechanism that may explain the surprisingly rapid evolution of HARs.

PrgmNr 3725 - Identification of ancestry specific health risks in a large cosmopolitan biobank

View session detail

Author Block: **C. Caggiano**[1], R. Shemirani[2], A. M. Chiu[3], R. Johnson[4], V. A. Arboleda[5], G. M. Belbin[6], N. Zaitlen[1]; [1]Dept. of Neurology, Univ. of California Los Angeles, Los Angeles, CA, [2]USC-ISI, Marina del Rey, CA, [3]Univ. of California Los Angeles, Los Angeles, [4]Dept. of Computer Sci., Univ. of California Los Angeles, Los Angeles, CA, [5]Univ California los Angeles, Los Angeles, CA, [6]Icahn Sch. of Med. at Mount Sinai, New York, NY
Disclosure Block: **C. Caggiano:** None.
INTRODUCTION: Genetic ancestry is a key factor that affects an individualâ□□s risk for disease. In clinical settings, however, genetic ancestry is underutilized. Instead, race and ethnicity are used as noisy proxies for ancestry, reducing the ability to identify any health disparities that may exist in genetic ancestry communities.
OBJECTIVES: In this work, we studied genetic ancestry in the context of a large cosmopolitan biobank, with the goal of identifying illnesses that may exist at an elevated prevalence in fine-scale ancestry groups. To do this, we used genotype and electronic health record (EHR) data from 28,000 individuals across the Los Angeles area.
METHODS: To assess genetic ancestry, we identified identical-by-descent (IBD) segments of biobank participants via iLASH. Using these IBD segments, we identified 180 genetic communities using the InfoMap community detection algorithm. Many of these communities are populations not well-studied in clinical human genetics, including groups characterized by the presence of Persian Jews, Armenians, and Ethiopians. After identifying these groups, we used deidentified EHR data to estimate population-specific disease risks.
RESULTS: In this analysis, we observed several well-known disease-ancestry associations, including an elevated risk for Crohnâ□□s Disease in Ashkenazi Jews versus non-Jewish Europeans (logistic regression p=3.82 x 10-8, odds-ratio: 2.0 95% CI: 0.60-3.40), along with novel associations in many non-European communities, such as an increased risk for sleep apnea in Persians (logistic regression p=8.43 x 10-4; odds-ratio: 2.5, 95% CI: 0.80-4.2).
CONCLUSIONS: These results demonstrate the value of using genetic ancestry in precision medicine initiatives, especially for understudied populations.

PrgmNr 3726 - Inferring negative natural selection at short tandem repeats in the human genome

View session detail

Author Block: **B. Huang**[1], K. E. Lohmueller[2,3], M. Gymrek[4,5]; [1]Dept. of Bioengineering, Univ. of California San Diego, La Jolla, CA, [2]Univ. of California Los Angeles, Los Angeles, CA, [3]Dept. of Human Genetics, David Geffen Sch. of Med., Univ. of California Los Angeles, Los Angeles, CA, [4]Dept. of Computer Sci. and Engineering, Univ. of California San Diego, La Jolla, CA, [5]Dept. of Med., Univ. of California San Diego, La Jolla, CA

Disclosure Block: **B. Huang:** None.

Short tandem repeats (STRs) are DNA sequences consisting of repeated 1-6 base pair motifs that represent about 1.6 million loci in the human genome. Due to their high prevalence in the genome and rapid mutation rates, variation in copy number at STRs represents a large portion of human genetic variation. Since STRs play an important role in gene expression and contribute to a wide range of human traits and disorders ranging from height and intelligence to cancer and Mendelian diseases, mutations at some STRs are likely to experience natural selection. Understanding these fitness effects can provide insights into the functional importance of particular classes of repeats. Recently, we introduced SISTR (Selection Inference at Short Tandem Repeats), a population genetics framework that estimates selection coefficients (*s*) at individual STRs. Although SISTR performs well on STRs with rapid mutation rates, the method is underpowered to detect selection at loci with low mutation rates or loci under modest selection. To overcome these limitations, we present SISTR2, an extension of SISTR that jointly estimates the distribution of *s* across a set of loci. SISTR2 assumes *s* for each STR is drawn from a gamma distribution and infers the parameters of this distribution, using approximate Bayesian computation and the set of allele frequency distributions for each locus. After validating SISTR2 on simulated datasets, we estimated the distribution of selection coefficients for different STR classes. As inputs, we used STR allele frequencies obtained from European individuals from the 1000 Genomes Project. We found that coding loci are under stronger negative selection than non-coding loci as expected. Furthermore, we performed a motif analysis and found that even among STRs with the same repeat unit length, there is notable variation in negative selection across different repeat unit sequences. For example, for dinucleotides, AT repeats (which are enriched in common human transposable element sequences and thus may be under decreased negative selection) appear more neutral than AC or AG repeats. For both trinucleotides and tetranucleotides, A_nG_n type repeats (e.g. AAG/AAAG/AAGG) generally appear under less selection. This is consistent with a previous observation that these repeats are most likely to expand in healthy controls, which may be caused by unstable base-stacking interactions at these STRs. Overall, a more accurate and nuanced understanding of selection coefficients at STRs of different repeat classes will be important for future studies of variation at STRs and their role in evolution and disease.

PrgmNr 3727 - Inferring population structure in biobank-scale genomic data

View session detail

Author Block: **A. M. Chiu**[1], E. K. Molloy[1,2], Z. Tan[3], A. Talwalkar[4], S. Sankararaman[1]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Univ. of Maryland, Coll. Park, College Park, MD, [3]Facebook, Inc., Menlo Park, CA, [4]Carnegie Mellon Univ., Pittsburgh, CA
Disclosure Block: **A.M. Chiu:** None.
Inference of population structure is a central problem in human genetics with applications ranging from understanding of human history to population stratification correction in genome-wide association studies. Approaches to population structure inference typically formalize the problem as one of estimating admixture proportions of each individual and ancestral population allele frequencies given genetic variation data. The growth of repositories of genetic variation data over large numbers of individuals has opened up the possibility of inferring population structure at increasingly finer resolution. This development has necessitated methods that can be applied to large-scale datasets with reasonable computational requirements. Thus, we developed SCOPE (SCalable pOPulation structure inferencE) - a scalable method capable of inferring population structure on biobank-scale data. SCOPE utilizes a previously proposed likelihood-free framework that estimates the individual allele frequency matrix through a statistical technique known as latent subspace estimation followed before decomposing it into ancestral allele frequencies and admixture proportions. SCOPE uses two ideas to substantially improve the scalability of this approach. First, SCOPE uses randomized eigendecomposition to efficiently estimate the latent subspace. Specifically, SCOPE avoids the need to form matrices that are expensive to compute or require substantial memory, instead working directly with the input genotypes. Second, SCOPE leverages the repeated multiplications of the genotype matrix and uses the Mailman algorithm for fast multiplication of the genotype matrix. We benchmarked the accuracy and efficiency of SCOPE on simulated and real datasets. In simulations, SCOPE obtains accuracy comparable to existing methods while being 3 to 1,800 times faster. Relative to the previous state-of-the-art scalable method (TeraStructure), SCOPE is 3 to 144 times faster. SCOPE can estimate population structure in about a day for a dataset consisting of one million individuals and SNPs whereas TeraStructure, is extrapolated to require approximately 20 days. We used SCOPE to infer continental ancestry proportions (four ancestry groups) on the UK Biobank dataset (488,363 individuals and 569,346 SNPs) in about a day. We find that the inferred continental ancestry proportions are highly concordant with self-reported race and ethnicity. SCOPE additionally can be applied in a supervised setting. Given allele frequencies from reference populations, SCOPE can estimate admixture proportions corresponding to the reference populations, to enable greater interpretability.

PrgmNr 3728 - Population level study of tandem repeats using an ensemble call-set

View session detail

Author Block: **N. Mousavi**, N. Ma, H. Ziaei Jam, B. Huang, M. Gymrek; UC San Diego, San Diego, CA
Disclosure Block: **N. Mousavi:** None.
Tandem repeats (TRs) are a class of highly variant genetic mutations characterized by motifs of 1-20 base pairs repeating in tandem. TRs form more than 3% of the human genome and have been associated with Mendelian disorders such as Huntingtonâ□□s disease and Fragile X syndrome. Furthermore, studies have linked TRs to complex traits such as gene expression. However, due to complexities in sequencing, assembly, and genotyping, TRs have been a relatively understudied class of genetic variants. Recently several tools (GangSTR, HipSTR, ExpansionHunter, AdVNTR, and others) have been developed for genome-wide and targeted genotyping of TRs from short-read sequencing data. While these methods open the doors to large-scale studies of these complex genetic variants, each method has different design principles that can lead to biased call sets. To address this issue, we have developed a graph-based method to create an ensemble TR call-set from the outputs of TR variant calling tools. Our method utilizes statistical measures of accuracy at each locus alongside knowledge on biases of each method to create a unified and unbiased output call-set. We applied this approach to the 2504 samples from the 1000 Genomes project and created a population-level unified ensemble TR call-set. We experimentally validate our ensemble call-set via fragment analysis at a subset of loci and samples and measure Mendelian error rate using 698 additional related samples. We calculate allele frequencies of TRs across different populations and find TRs with highly population-specific alleles. We further measure negative selection at different TR loci using SISTR and identify specific loci that are likely to be under selection pressure. Finally, we identify and characterize TR expansions polymorphic throughout different populations. Our results create an unbiased TR call-set on a large-scale dataset and shed light on the population-level variability of TRs.

PrgmNr 3729 - Reconstructing spatio-temporal patterns of admixture in human history using present-day and ancient genomes

View session detail

Author Block: **M. Chintalapati**[1], N. J. Patterson[2], N. Alex[1], P. Moorjani[1]; [1]Univ. of California, Berkeley, CA, [2]Broad Inst MIT & Harvard, Cambridge, MA
Disclosure Block: **M. Chintalapati:** None.
Recent studies of contemporary and ancient human populations has shown that gene flow or admixture has been pervasive throughout human history. Understanding the timing and signatures of admixture is important for studying the evolutionary and functional impact of admixture, as well as uncovering the historical context of the mixture. With the availability of a large number of present-day and ancient genomes, it is now possible to characterize the timing of demographic changes with unprecedented resolution. While a number of methods exist for characterizing population mixture in contemporary populations, they are not applicable to sparse data available from ancient DNA specimens (with low coverage and large proportions of missing genotypes). To address this need, we developed *DATES* (Distribution of Ancestry Tracts of Evolutionary Signals) that leverages ancestry covariance patterns in a single diploid genome to infer timing of admixture. By simulations, we show that *DATES* provides reliable results under a range of scenarios, including for cases with only a single admixed individual, admixed genomes with low coverage and admixed genomes with large amounts of missing data. We applied *DATES* to reconstruct the timing of population admixture using Human Origins dataset with over 7,000 present-day and 5,000 ancient human genomes, including the spread of Mongol Empire expansion and formation of present-day Japanese ancestry. Using time-transect data from diverse geographic regions and cultures in Europe, we perform a fine-scale reconstruction of the mixture of ancient hunter-gatherers, Neolithic farmers and Steppe-pastoralists uncovering the dynamics of Neolithic farming and the spread of Indo-European languages in Europe. Our analyses highlight the power of genomic dating methods to elucidate the legacy of human migrations and provide complementary insights to archaeological and linguistic evidence.

PrgmNr 3730 - Reconstructing the 3D chromatin organization of archaic hominins reveals that genome folding shaped human-Neanderthal phenotypic divergence and introgression

View session detail

Author Block: **J. A. Capra**[1], E. McArthur[2], D. C. Rinker[2], E. Gilbertson[1], G. Fudenberg[1], M. Pittman[1], K. Keough[3], K. S. Pollard[4]; [1]Univ. of California San Francisco, San Francisco, CA, [2]Vanderbilt Univ., Nashville, TN, [3]Gladstone Inst.s, San Francisco, CA, [4]Univ California San Francisco, San Francisco, CA
Disclosure Block: **J.A. Capra:** None.
Introduction: Modification of gene regulation was a driving force in the divergence of modern humans and archaic hominins. While previous investigation has focused on changes in cis-regulatory elements, the three-dimensional (3D) organization of the genome plays a critical role in regulating gene expression by facilitating and insulating enhancer-promoter interactions. However, the role of 3D genome organization changes in recent human evolution has not been explored because the degradation of ancient samples does not permit experimental interrogation of archaic hominin 3D genome folding.
Results: We address this gap by applying novel deep learning methods for inferring 3D genome organization from DNA sequence patterns to Neanderthal, Denisovan, and diverse modern human genomes. Using the resulting genome-wide 3D genome folding maps, we find that Neanderthals have more predicted 3D genome similarity to humans than expected based on sequence similarity. This suggests that pressure to maintain 3D conformation has constrained sequence divergence in recent hominin evolution. Although humans and Neanderthals have broadly similar 3D genome structures, we identify many archaic-specific and modern human-specific 3D patterns (e.g. chromatin loops). Thus, Neanderthal introgression had the potential to impart divergent 3D genome folding to Eurasians. Evaluating the legacy of introgression on human 3D genome organization, we identify substantial changes in 3D folding patterns caused by introgressed variants that also associate with traits relevant to human-Neanderthal differences, including height, fat distribution, and blood pressure. Additionally, we discover that tolerance to 3D genome variation in humans constrained Neanderthal introgression: regions more tolerant of 3D variation in modern Africans are enriched for introgression in modern Eurasians.
Conclusion: In summary, our approach opens a window into previously unobservable molecular mechanisms underlying how genetic differences lead to phenotypic divergence between modern and archaic hominins by applying deep learning to predict archaic 3D genome folding.

PrgmNr 3731 - The mutational dynamics of short tandem repeats in large, multigenerational families

View session detail

Author Block: **C. Steely**[1], W. Watkins[2], L. M. Baird[1], L. B. Jorde[3]; [1]Univ. of Utah, Sch. of Med., Salt Lake City, UT, [2]Univ. of Utah, Salt Lake City, UT, [3]Univ of Utah, Salt Lake City, UT
Disclosure Block: **C. Steely:** None.
Short tandem repeats (STRs) are tandemly repeated sequences of 1-6 bp motifs. These sequences compose approximately 3% of the genome, and mutations at STR loci have been linked to dozens of human diseases including amyotrophic lateral sclerosis, Friedreich ataxia, Huntington disease, and fragile X syndrome. Improving our understanding of these mutations would increase our knowledge of the mutational dynamics of the genome and may uncover additional loci that have a role in causing disease. Here, to estimate the genome-wide pattern of mutations at STR loci, we examined blood-derived whole-genome sequencing data for 544 individuals from 29 three-generation CEPH pedigrees. These pedigrees contain both sets of grandparents (generation 1), the parents (generation 2), and an average of 9 grandchildren (generation 3) per family. We use HipSTR to identify *de novo* STR mutations in the 2nd generation of these pedigrees. Analyzing approximately 1.6 million STR loci, we find an average *de novo* STR mutation rate of 5.24*10-5 mutations per locus per generation. We find that perfect repeats mutate ~3x more quickly than imperfect repeats and *de novo* STRs intersect with transposable elements. Approximately 30% of these *de novo* STRs are found in *Alu* elements, even though these elements compose only 11% of the genome. Another ~10% are found in LINE-1 insertions, which compose 17% of the genome. Phasing these *de novo* mutations to the parent of origin shows parental transmission biases that vary among families. We estimate the average number of *de novo* genome-wide STR mutations per individual to be ~85, which is coincidentally similar to the average number of observed *de novo* single nucleotide variants.

PrgmNr 3732 - The waiting distance distribution in ancestral recombination graphs

View session detail

Author Block: **Y. Deng**; UC Berkeley, Berkeley, CA
Disclosure Block: **Y. Deng:** None.
The ancestral recombination graph (ARG) contains the full genealogical information of the sample, and many population genetic inference problems can be solved using inferred or sampled ARGs. In particular, the waiting distance between tree changes along the genome can be used to make inference about the distribution and evolution of recombination rates. To this end, we here derive an analytic expression for the distribution of waiting distances between tree changes under the sequentially Markovian coalescent model and obtain an accurate approximation to the distribution of waiting distances for topology changes. We use these results to show that some of the recently proposed methods for inferring sequences of trees along the genome provide strongly biased distributions of waiting distances. In addition, we provide a correction to an undercounting problem facing all available ARG inference methods, thereby facilitating the use of ARG inference methods to estimate temporal changes in the recombination rate.

PrgmNr 3733 - A retrospective study of health issues being discussed in a prenatal setting with female carriers of fragile X syndrome

View session detail

Author Block: **F. Ellahy**[1], S. Chen[2], T. Giordonello[2], M. Zak[1]; [1]Long Island Univ. Post, Brookville, NY, [2]Northwell Hlth., Great Neck, NY
Disclosure Block: **F. Ellahy:** None.
Fragile X syndrome is a complex genetic condition that has health implications for carriers as well as affected individuals. Over the past decade, a multitude of scientific literature has been published demonstrating a correlation between the premutation allele and increased incidences of various physical and mental issues beyond fragile X-associated tremor and ataxia syndrome (FXTAS) and premature ovarian failure (POF). Clinicians counseling these patients should be updated about this recent literature to provide patients with the most accurate information to make informed choices about their health and reproduction. As there is no literature published regarding what health issues are being discussed with fragile X carriers in genetic counseling sessions, one of this studyâ□□s goals is to shed light on this area. Furthermore, by analyzing the trends over the years, this study also aims to determine if the number of topics discussed with these patients has changed. In this retrospective chart review, 107 patient records spanning the years 2012 to 2020 were obtained. Since fragile X carrier screening is usually performed in a prenatal setting, this study obtained its patient population from four prenatal clinics. The study population included female intermediate or premutation carriers between the ages of 18 to 50 years. This study found that health issues strongly associated with premutation alleles, such as FXTAS and POF, were the most frequently discussed topics (44.9% of N). Other conditions with less conclusive evidence regarding association with the premutation allele, such as learning disabilities and psychological disorders (autism, anxiety, and depression), were less frequently discussed (~3% of N) but are starting to be included in the discussion. This study found an overall increase in the average number of topics discussed with carriers over the years 2012-2020 (premutation carriers: 3.00 to 4.86, intermediate carriers: 0 to 2.25). The findings from this study demonstrate that conversation regarding emerging health problems associated with premutation carriers is starting to take place. The results also suggest that clinicians are staying updated with the literature and adapting their counseling to include this new research in their discussions. This studyâ□□s findings can be used as a resource by clinicians and genetic counseling programs to inform their practices by identifying the topics currently being discussed with this patient population and to foster a discussion about improving the quality and content of counseling this patient cohort.

PrgmNr 3734 - Analysis of Financial Barriers Experienced by Prospective Genetic Counseling Students

View session detail

Author Block: **D. Lee**[1], J. Platt[2,3], P. Flodman[1], K. Singh[1], F. Quintero-Rivera[1,4]; [1]Div. of Genetic and Genomic Med., Dept. of Pediatrics, Univ. of California, Irvine, Irvine, CA, [2]Stanford Ctr. for Inherited Cardiovascular Disease, Stanford Hlth., Palo Alto, CA, [3]Div. of Med. Genetics, Dept. of Pediatrics, Stanford Univ. Sch. of Med., Palo Alto, CA, [4]Dept. of Pathology and Lab. Med., Univ. of California Irvine, Irvine, CA

Disclosure Block: **D. Lee:** None.

Background: Genetic counseling (GC) graduate program applicants are burdened with application fees and the â□□hidden costsâ□□ of applying, including graduate exam fees, prerequisite coursework, travel expenses for interviews, and time off from work to obtain relevant volunteer experiences. These costs can become prohibitively expensive for many applicants, especially those with fewer resources. Furthermore, the high costs of applying can become a barrier to diversify the workforce.

Objectives: This study aimed to address the following: 1) What were the median application costs for prospective students applying to GC programs in the United States? 2) What aspects of the application process were most expensive? 3) Were there differences between individuals of historically underrepresented racial and ethnic backgrounds in medicine (hURM) and non-underrepresented applicants with respect to total application costs, accrued volunteer hours, parental education, and familial financial assistance?

Methods: A survey was developed to capture demographic information, application history, application and preparation expenses, time volunteering, and financial resources. A total of 383 responses were analyzed.

Results: Median total application costs for respondents who attempted one application cycle were $2,634 (n = 264, range: $202 - $25,693). For those who attempted twice, median total costs were $4,762 (n = 84, range: $909 - $24,206). Interview-related items had the highest median cost (one application cycle: $879, range: $0 - $6,007; two or more application cycles: $1,310, range: $0 - $7,307). Among those who applied to more than one cycle, hURM respondents (n=19) had higher median total costs ($6,713 versus $4,762, p = 0.03) and lower median total volunteer hours (246 versus 381 hours, p =0.03) than those of non-underrepresented individuals (n = 100). Additionally, parental education level differed (p = 0.04) between the two groups. Higher parental education level was correlated with a higher percentage of familial financial support (2% median financial contribution from parents with a high school diploma or associate degree versus 60% from parents who both have advanced degrees, p = 0.0009).

Conclusion: This study demonstrates the significant costs associated with the GC application process. Differences in costs and resources were also observed between applicants of differing socioeconomic status. Stakeholders within the profession should implement strategies to reduce financial barriers and the resulting inequities in the application process, which will improve access to GC graduate programs and enhance efforts to diversify the workforce.

PrgmNr 3735 - Attitudes towards polygenic risk testing for glaucoma in an Australian population

View session detail

Author Block: **G. Hollitt**[1], O. M. Siggs[1], B. Ridge[1], M. C. Keane[1], D. A. Mackey[2], S. Macgregor[3], A. W. Hewitt[4], J. E. Craig[1]; [1]Dept. of Ophthalmology Flinders Univ., BEDFORD PARK, Australia, [2]Lions Eye Inst., Nedlands, Australia, [3]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [4]Ctr Eye Res Aust, Brunswick, Australia

Disclosure Block: **G. Hollitt:** Major Stockholder/Ownership Interest; S MacGregor, AW Hewitt and JE Craig are listed as co-inventors on a patent application for the use of genetic risk scores to determine risk and guide treatment for glaucoma.

Despite the significant progress made in genetic risk prediction, critical gaps in knowledge pertinent to understanding barriers to the implementation of polygenic risk scores (PRS) testing persist. We performed a cross-sectional, questionnaire-based study to better understand the attitudes of individuals with and without glaucoma toward PRS for glaucoma. As the leading cause of irreversible vision loss in the world, with recognised complex heritability, few environmental risk factors, and high treatability to prevent blindness, PRS testing has strong clinical utility for glaucoma. We surveyed 1169 individuals with glaucoma and 418 individuals without glaucoma to evaluate their attitude toward polygenic risk testing for glaucoma. Those without glaucoma included those with a first-degree relative with diagnosed glaucoma, those attending an optometrist without glaucomatous features, and unselected members of the community. We assessed several factors affecting interest in testing using multivariate regression analysis. Our results showed strong interest in the test, with 69.4% of individuals with glaucoma and 71.3% of individuals without glaucoma indicating a keenness in testing. Among individuals with glaucoma, interest was seen in those from an urban area (OR 1.696, 95% CI (1.154-2.494), p=0.007), those who perceived their risk of developing glaucoma as higher (OR 2.053, 95% CI (1.280-3.293), p=0.003), and those who were worried about developing glaucoma (OR 2.068, 95% CI (1.270-3.369), p=0.004). In individuals without glaucoma, interest was seen in those who perceived their risk of developing glaucoma was higher (OR 14.579, 95% CI (1.146-185.5), p=0.039), those who were worried about developing glaucoma (OR 4.371, 95% CI (2.317-8.246), p

PrgmNr 3736 - Authorship Guidelines and Implementation: Promoting Equity and Inclusion

View session detail

Author Block: **H. Lewis**[1], S. S. Lee[2], K. A. Goddard[3], B. S. Wilfond[4], B. B. Biesecker[5]; [1]Seattle Children's Hosp., Seattle, WA, [2]Columbia Univ., New York, NY, [3]Kaiser Permanente Northwest, Portland, OR, [4]Seattle Children's Res. Inst., Seattle, WA, [5]RTI Intl., Washington, DC
Disclosure Block: **H. Lewis:** None.
Prioritizing equity in authorship is a challenge for large research teams. Navigating across institutions, disciplines, career stage, and team size contributes to the complexities. The Cancer Health Assessments Reaching Many (CHARM) study is exemplary of this layered challenge. CHARM integrates a focus on Team Science that aims to elevate the effectiveness of research teams by prioritizing diversity and inclusion through collaboration, conflict resolution, and effective communication. This approach facilitated a process to gather reflections and feedback from all CHARM investigators on prior or existing experiences with authorship that identified a need for a more equitable process. Using a Team Science approach, standard authorship guidelines for CHARM were developed to strive for better equity and inclusion. CHARM also developed an electronic matrix to track the number of manuscripts, team member authors, and distribution of authorship opportunities. Using this matrix, the CHARM team can further its goal of inclusion by identifying opportunities for those who have had fewer authorship opportunities, through mentorship of junior researchers as first authors by experienced senior authors. By clarifying authorship and writing group roles in guidelines, expectations of the level of involvement in a manuscript can be clarified and differing views on authorship position can be avoided. While CHARM authorship guidelines are specific to the team, the development process can be appropriated within other large research teams. The process included assessing authorship expectations and norms to reveal existing assumptions; soliciting specific issues and concerns with existing or prior authorship experiences; using facilitated small group discussions to illuminate optimal outcomes; and deliberating on guideline priorities within a small working group that can be taken to the team for consensus. Team leaders can configure these priorities into a fair and equitable process to be implemented across the team following final discussions and input from all team members.

PrgmNr 3737 - Clinical Sequencing Evidence-Generating Research (CSER) consortium Diversity and Inclusion Statement: development, revision, and adoption

View session detail

Author Block: **S. M. Fullerton**[1], S. J. Knight[2], A. M. Gutierrez[3], B. B. Biesecker[4], K. A. Goddard[5], G. LaMoure[6], M. A. Majumder[3], P. Murali[1], S. Outram[7], A. B. Popejoy[8,9], K. Renna[6], D. J. Kaufman[6], L. A. Hindorff[10]; [1]Univ. of Washington, Seattle, WA, [2]Univ. of Utah, Salt Lake City, UT, [3]Baylor Coll. of Med., Houston, TX, [4]RTI Intl., Washington, DC, [5]Kaiser Permanente Northwest, Portland, OR, [6]Natl. Human Genome Res. Inst., Bethesda, MD, [7]UCSF, San Francisco, CA, [8]Stanford Univ., Stanford, CA, [9]Univ. of California, Davis Sch. of Med., Davis, CA, [10]NIH, Bethesda, MD
Disclosure Block: **S.M. Fullerton:** None.
Background: The Clinical Sequencing Evidence-Generating Research (CSER) Consortium was initiated in 2017 with the aim of investigating the effectiveness of integrating genomic (exome or genome) sequencing into the clinical care of medically underserved individuals across diverse health care settings and disease states. At the recommendation of participant stakeholders, and in the wake of broader discussion about the impact of systemic racism and other forms of discrimination in many contexts including biomedical research, the consortium agreed to adopt a formal Diversity and Inclusion Statement. Here we describe the final statement ratified by CSER and the process by which the statement was developed, reviewed, revised, and adopted for use in consortium publications and presentations. **Methods**: The suggestion to adopt a consortium-wide Diversity and Inclusion statement was brought to a Steering Committee meeting in the summer of 2020. Following discussion with CSER principal investigators and workgroup chairs, an *ad hoc* group of interested CSER investigators, with representation from all eight CSER sites and the NHGRI, was convened to draft a statement for adoption and use by the consortium. A preliminary draft was presented to scientific investigators and a group of patient-participant stakeholders for feedback. The statement was revised further and circulated to CSER sites for final review before formal ratification at a Steering Committee meeting. The statement was also translated into Spanish to make it accessible to Spanish-preferring stakeholders. **Results**: The iterative drafting, review, and revision process resulted in a brief (approximately 100 word) plain language statement that describes CSERâ□□s commitment to understanding how genes impact disease and reducing barriers to genetic services among people who are discriminated against, marginalized, and medically underserved. The statement addresses stakeholder engagement and workforce diversity, as well as responsible analysis and reporting. While acknowledging that much work needs to be done to achieve a just and equitable health care future, the statement emphasizes CSERâ□□s commitment to creating genomic knowledge that will benefit people of all backgrounds. **Conclusions**: Although intended primarily to guide CSER investigators and stakeholders in the final stage of their collective work together, we hope that the statement can inform the work of the clinical translational genomics research community moving forward.

PrgmNr 3738 - Cloud-based biomedical data storage and analysis for genomic research: landscape analysis of data governance in emerging NIH-supported platforms

View session detail

Author Block: **S. C. Nelson**[1], J. Dahlquist[2], S. M. Fullerton[2]; [1]Biostatistics, Univ. of Washington, Seattle, WA, [2]Bioethics & Humanities, Univ. of Washington, Seattle, WA
Disclosure Block: **S.C. Nelson:** None.
Background: Genomic research initiatives and consortia are generating individual-level genomic, environmental, and linked phenotypic and/or health outcome data at an unprecedented pace and scale, driving innovation in the storage, sharing, and analysis of such data. At the forefront of new data sharing mechanisms are institutional and, increasingly national, cloud-based computing and storage platforms designed to support efficient and scalable data processing, analysis, and sharing. In parallel, efforts to streamline and even partially automate data access are being pursued. These new data sharing platforms and procedures are designed to facilitate collaboration and maximize the scientific utility of costly-to-generate genomic and linked clinical data. Yet the implications of current access procedures and data protections standards for trustworthy stewardship remain underexamined by ethicists and policymakers. **Methods**: To inform policy development for the responsible governance of cloud-based biomedical data and analysis, we conducted a landscape analysis of data governance practices in three newly developed NIH-supported cloud platforms, the NHLBI BioData Catalyst, the NHGRI AnVIL (Analysis, Visualization, and Informatics Lab-space), and the All of Us Research Hub, as well as two predecessor data sharing platforms, the NCBI Database of Genotypes and Phenotypes (dbGaP) and the NCI Genomic Data Commons. **Results**: In this presentation we report key initial findings from that analysis, based in a content analysis of platform documentation, participant observation of platform developer discussions, and key informant interviews with platform developers and related leadership. **Conclusions**: Our review of current platform governance policies and procedures identifies numerous tradeoffs with potential implications for both participant and public trust in precision medicine research.

PrgmNr 3739 - Conducting clinical genomic medicine research during the COVID-19 pandemic: Lessons learned from the CSER consortium experience

View session detail

Author Block: **S. Kraft**[1,2], J. T. Bensen[3], K. E. Donohue Bonini[4], L. A. Hindorff[5], D. J. Kaufman[6], K. E. Renna[7], J. O. Robinson[8], H. Russell[9], N. Sahin-Hodoglugil[10], M. Waltz[11], J. M. Zepp[12], S. J. Knight[13]; [1]Seattle Children's Res. Inst., Seattle, WA, [2]Univ. of Washington Sch. of Med., Seattle, WA, [3]Univ North Carolina at Chaper Hill, Chapel Hill, NC, [4]Inst. for Genomic Hlth., Icahn Sch. of Med. at Mt. Sinai, New York City, NY, [5]NIH, Bethesda, MD, [6]NIH - NHGRI, Bethesda, MD, [7]Div. of Genomic Med., Natl. Human Genome Res. Inst., Bethesda, MD, [8]Baylor Coll. Med., Houston, TX, [9]Dept. of Pediatrics Oncology, Baylor Coll. of Med., Houston, TX, [10]Inst. for Human Genetics, Univ. of California-San Francisco, San Francisco, CA, [11]Univ. of North Carolina, Chapel Hill, Chapel Hill, NC, [12]Dept. of Translational and Applied Genomics, Ctr. for Hlth.Res., Kaiser Permanente Northwest, Portland, OR, [13]Univ. of Utah, Salt Lake City, UT

Disclosure Block: **S. Kraft:** None.

The COVID-19 pandemic has led to significant changes in healthcare practices and priorities and therefore has had a major impact on genomic medicine research. Research studies have faced challenges stemming from a wide range of clinical and research factors including new procedures and policies for in-person visits, the rapid implementation of telehealth and other remote procedures, and changing priorities for participants, clinicians, and research team members. Further, multi-site projects and research consortia have had to navigate variation between institutional approaches. These challenges have affected all stages of genomic medicine researchâ□□from recruitment and consent through results disclosure and clinical follow-up. These challenges to research participation may be particularly burdensome to participants from historically marginalized racial, ethnic, and socioeconomic groups who been disproportionately impacted by COVID-19, continuing the pattern of long-standing barriers to participation in genomic research.

This presentation will describe the experience of six genomic medicine studies across ten states within the Clinical Sequencing Evidence-Generating Research (CSER) consortium as they navigated the COVID-19 pandemic, highlighting challenges faced, strategies for overcoming these challenges, and lessons learned that may inform future research operations. CSER is an NIH-funded research consortium studying the integration of genome sequencing into clinical care for diverse and medically underserved individuals. We interviewed and surveyed representatives of the six extramural CSER projects at multiple time points during 2020 and 2021. Using a descriptive content analysis approach, we identified two overarching areas in which projects faced particularly impactful challenges and saw opportunities for future improvement: (1) participant recruitment, enrollment, and engagement; and (2) the transition to telehealth and other remote research procedures. In this presentation, we will describe how projects navigated these challenges throughout the COVID-19 pandemic, with a focus on ethical considerations related to the evolving clinical-research boundary and access to genomic services for medically underserved patients. We will discuss key lessons learned to advance equitable genomic medicine research moving forward.

PrgmNr 3741 - Designing and launching a pilot race and genetics seminar series for public health genetics students

View session detail

Author Block: **P. Murali**, A. Khan; Univ. of Washington, Seattle, WA
Disclosure Block: **P. Murali:** None.
The field of genetics continues to grapple with issues of race and racism, but adequate exposure to these issues in graduate training programs is lacking. Concepts of race, ethnicity, and ancestry are discussed in human genetics coursework, but little attention is given to understanding the relevance, purpose, and context of their use. Awareness of these concepts is important to avoid reinforcing notions of racial essentialism. To address this, we designed a ten-week race and genetics seminar series for the Public Health Genetics (PHG) department at the University of Washington in order to provide a space for PHG students to engage with these complex topics.
In this seminar series, we examined historical and current perspectives on race, and its role in genetic and biomedical research, as well as in health inequities. We developed a framework for reflecting on PHG core coursework and making suggestions for areas of anti-racist intervention. Examples include addressing race, ethnicity, and ancestry in (1) genetic association methods for analysis of diverse populations in the field of genetic epidemiology; (2) policy and legal cases in the intersection of genetics and the law; and (3) methods of analysis and inference for drug development in different populations in the area of pharmacogenetics. Finally, we administered an exit survey to gain feedback on the utility of the seminar. Responses from the survey illustrated that participantsâ□□ perceptions about race, racism, and genetics changed as a result of participation in the seminar, and that they gained knowledge about these topic areas. Respondents also provided suggestions for how to incorporate concepts learned in the seminar into the broader PHG training program. Ultimately, this student-driven seminar series provided opportunities for participants to wrestle with complex and challenging concepts of race and genetics. Furthermore, this seminar can be incorporated into other genetics graduate programs, such as genetic counseling and medical genetics residency programs. Genetics researchers and professionals play a pivotal role in influencing the way concepts of race, genetics, and health are interpreted in research, medical care, and policy. Therefore, exposure to these concepts with an interdisciplinary approach during graduate training is an important step towards eliminating notions of racial essentialism and improving research methods, interpretation, and application.

PrgmNr 3742 - Ethnic Differences in the Frequency of Cancer Reported from Family Pedigrees in the Prenatal Genetic Counseling Setting

View session detail

Author Block: **A. Palacios**[1], P. Flodman[2], K. French[3], M. Smith[4]; [1]Dept Pediatrics, Divn of Genetic and Genomic Med., Univ California Irvine Med Ctr, Orange, CA, [2]Univ California Irvine Med Ctr, Orange, CA, [3]Dept Pediatrics, Divn of Genetics and Metabolism, Univ California Irvine Med Ctr, Orange, CA, [4]Univ California Irvine, Irvine, CA

Disclosure Block: **A. Palacios:** None.

This study analyzed if differences in cancer reporting exist between different ethnic groups when collecting family pedigrees in a prenatal genetic counseling setting. Data was collected from 446 prenatal charts at University of California, Irvine from January 1, 2015 - August 31, 2020. A total of 795 pedigrees meeting inclusion criteria (409 maternal pedigrees and 386 paternal pedigrees) were analyzed from four ethnic groups: non-Hispanic White, Hispanic/Latino(a), Asian, and African American/Black. The total number of first- and second-degree relatives and number of these relatives affected with cancer were calculated for each pedigree and analyzed using contingency tables, non-parametric tests, and Poisson regression. Cancer reporting in first- and second-degree relatives was the highest among the non-Hispanic White group. Reporting of a family history of cancer was lower in Hispanics, Asians, and African Americans. Ethnicity was a significant factor in predicting the number of relatives reported to have cancer in a Poisson regression model (controlling for the total number of relatives in the pedigree). The incidence of cancer reported in the pedigrees for Hispanics, African Americans, and Asians was 36.3%, 50.2%, and 65.5% (respectively) of the incidence seen in the non-Hispanic White pedigrees. The cancer reporting differences observed in the Asian pedigrees are similar to differences in population incidence as reported by the CDC, but the reporting in the Hispanic and African American pedigrees is less than would be expected based on population incidence. This suggests that cancer histories in some minority populations may be truncated. Genetic counselors, clinical geneticists, and Maternal Fetal Medicine providers should recognize that certain patient populations may be at risk for limited knowledge of a family cancer history. The study also identified that cancer reporting in the paternal family history was significantly increased when the father was present (p

PrgmNr 3743 - Exploring a role for genetic counselors in changing client health behavior: An Australasian focus group study

View session detail

Author Block: **E. Turbitt**[1], C. Jacobs[1], A. McEwen[1], L. Atkins[2]; [1]Univ. of Technology Sydney, Ultimo, Australia, [2]UCL Ctr. for Behaviour Change, London, United Kingdom
Disclosure Block: **E. Turbitt:** None.
Many cite undergoing genetic testing to gain information that will prompt a change in behavior, leading to better health outcomes. However, provision of genetic information alone has limited influence over health behavior. Genetic counselors possess relevant skills to facilitate health behavior change, though the extent to which they are ready and willing to be more involved in changing client health behavior is unknown.
We conducted a cross-sectional qualitative study to gain an in-depth understanding of genetic counselorsâ current practices and opinions about their involvement in changing client health behavior. Four virtual focus groups were conducted via Zoom video conferencing with genetic counselors (n=20) across Australasia. Evidence-based behavior change models (the Theoretical Domains Framework; TDF, and the Capability, Opportunity, Motivation -Behavior; COM-B) guided development of the focus group schedule, data analysis and interpretation. Themes were identified from verbatim transcripts using thematic analysis.
Our findings revealed three key target behaviors that genetic counselors hope clients would change following genetic counseling: (1) attend recommended screening/health appointments, (2) access information and support, and (3) share accurate information with relevant family members. We further identified genetic counselorsâ behaviors that aim to facilitate clientsâ behavior change such as activating screening/health appointments, providing access to relevant support, and equipping clients with skills for family communication. Influencers and barriers to genetic counselorsâ behaviors included having local knowledge and connections and beliefs about a tension between non-directiveness and behavior change. Influencers and barriers to changing behaviors included clientsâ willingness to act and having resources to carry out behaviors.
Genetic counselors may be interested and willing to have a role in changing client health behaviors. Key influencers and barriers identified in our study will be targeted in an intervention to enhance awareness and understanding of behavior change techniques among genetic counselors. Changing client health behaviors is a key factor to enabling the transformative role of genomics in disease prevention and management.

PrgmNr 3744 - Impact of the COVID-19 pandemic on the training and well-being of Medical Genetics and Genomics trainees

View session detail

Author Block: **J. Chenbhanich**[1], A. Tam[1], A. M. Slavotinek[2]; [1]Univ. of California, San Francisco, San Francisco, CA, [2]Univ California San Francisco, San Francisco, CA
Disclosure Block: **J. Chenbhanich:** None.
Background: The COVID-19 pandemic has an unprecedented impact on healthcare systems and graduate medical education. We sought to understand how the COVID-19 pandemic with telehealth and virtual education has affected the well-being, clinical training, and medical education for clinical trainees in the Medical Genetics and Genomics Residency and Fellowship programs. **Methods**: All clinical trainees in the American Board of Medical Genetics and Genomics (ABMGG)-accredited Clinical Genetics and Genomics, and Clinical Biochemical Genetics training programs in the United States were emailed an anonymous, voluntary, 30-question online survey on March 22, 2021. The survey window closed on April 30, 2021. **Results**: 31 out of a possible 174 trainees (18%) completed the survey. With regard to well-being, trainees reported increased anxiety (18/31), increased depression (10/31), worsening work-life balance (3/31) and worsening physical health (fitness, sleep, and/or nutrition; 13/31). There was an approximate increased utilization of telehealth in outpatient clinical encounters from 3% pre-pandemic to 67% during the pandemic; separately, an approximate increase in telehealth utilization in inpatient clinical encounters from 0% pre-pandemic to 29% during the pandemic. The most commonly reported challenges in telehealth utilization included inadequate physical examination (n=26), technical problems during visits (n=25), and limited access only to patients with internet access (n=20). Eight trainees would like to continue the same amount of telehealth as part of clinical training after the pandemic, whereas 20 trainees believed that the COVID-19 pandemic has negatively impacted overall clinical training. Twenty-five trainees agreed that the pandemic has had a negative impact on education opportunities not provided by their training programs, such as conferences. **Conclusions**: This is the first formal evaluation of the COVID-19 pandemic impact on trainees who are in the Medical Genetics and Genomics clinical training programs to our knowledge. While the sample size is relatively small, we illustrate that the COVID-19 pandemic has negatively impacted many trainees in Medical Genetics and Genomics, with effects on their well-being, clinical training, and education. Telehealth has been increasingly used in both inpatient and outpatient encounters with some challenges. Further studies are needed to optimally integrate what we have learned about virtual training, including telehealth and virtual meetings, into the education of medical genetics in the post-pandemic era.

PrgmNr 3745 - Medical genetics among Internal Medicine residents

View session detail

Author Block: **M. Jose**, A. Pathak; Saint Peter's Univ. Hosp., New Brunswick, NJ
Disclosure Block: **M. Jose:** None.
Introduction - With rapid advances in medicine and patients with genetic conditions surviving to adulthood, medical genetics has expanded to all age groups. A recent study showed that we have two clinical geneticists per one million population. A major step in mitigating the workforce shortage includes filling the void at the training level. Hence, we assessed the awareness of Medical Genetics training among Internal Medicine (IM) residents in a community program. **Objective** - Our study aimed to assess 1) the awareness of medical genetics among IM residents 2) the enthusiasm of IM residents to know more about the field and 3) assess the areas that residents were interested in exploring. **Method** - All internal medicine residents (49 residents except for the two residents who are the authors of the study) at Saint Peters University Hospital in NJ were eligible for the study. The survey was distributed through social media platform using Survey Monkey. The survey consisted of 10 questions. Participation was voluntary. **Results** - 81.6% (40/49) of the residents responded to the survey. 70% of the IM residents identified the primary role of medical geneticist as clinical diagnosis and genetic counseling while 15% identified it as purely genetic counseling and 5% labeled it as research-related. 12.5% of the residents were not aware that training in medical genetics exists. 50% were not aware that a medical geneticist was available at their training site. 10% wanted to refer their patient to a geneticist but did not know how to proceed with it and 12.5% were not aware of the clinical criteria for referral. 17.5% wanted to order a genetic test but did not know how to proceed with it. 30% had interpreted a genetic test result to their patient and 50% of those stated, they wanted to do it better. 80% of the residents wanted to know more about the specialty. 75% of the residents stated that their interest in pursuing medical genetics will depend on knowing more about the specialty and the training. 85% of the IM residents wanted a medical genetics curriculum introduced to their course. **Conclusion** - Our analysis reflects the limited inclusion of medical genetics in the Internal Medicine training curriculum despite the growing demand in the field. More than three-quarters of the residents who participated in the survey wanted a genetics curriculum incorporated into their training. Increasing the awareness among trainees will help with the staffing shortage and also help medical geneticists have a healthy partnership with primary care providers and thus ensure timely referral and early diagnosis.

PrgmNr 3746 - Perspectives of Genetics Professionals on the Reporting of Variants of Uncertain Significance (VUS): Should they always be returned?

View session detail

Author Block: **E. D. Esplin**[1], M. A. Fox[2], S. C. Saitta[3], A. B. Santani[4], S. Aradhya[1], J. Bernstein[5], R. L. Nussbaum[1]; [1]Invitae, San Francisco, CA, [2]Invitae, Encino, CA, [3]David Geffen Sch. of Med. at UCLA, Los Angeles, CA, [4]Veritas Genetics, Philadelphia, PA, [5]Stanford Univ Sch Med, Stanford, CA
Disclosure Block: **E.D. Esplin:** Major Stockholder/Ownership Interest; Invitae. Salary/Employment; Invitae.
Background. Variants of uncertain significance (VUS) are common. With increasing use of multi-gene panels (MGP), the number of VUS grows. Not all VUS have the same degree of uncertainty, presenting challenges to non-genetics healthcare providers (HCPs), patients and families. We report the results of a survey assessing attitudes of genetics professionals on the return of VUS results from MGP, emphasizing cancer genetics professionals in academic settings.
Methods. We designed a survey regarding VUS in MGP and whole exome sequencing (WES); this analysis focused on responses in the context of VUS in MGP. The survey was sent to ASHG and NSGC members to query those having experience with MGP, 2711 confirmed receipt. Initial analysis: 378 respondents (14%), 253 completed the survey (9%); 66% genetic counselors, 21% clinical geneticists, among others; 60% academic, 31% in a private clinical or commercial laboratory setting. 40% work primarily in cancer genetics, 13% in pediatric genetics, 11% adult genetics, skewing the respondents towards those who assess VUS for dominant/incompletely penetrant conditions.
Results. Overall, 47% (total n=167) agree that VUS can be useful for patient care with 30% in disagreement. In the last year, most respondents (75%) received a VUS result they considered clinically significant (prompted family segregation studies, was in trans with P/LP variant in an autosomal recessive condition, etc.), with 67% receiving **Conclusions**. While VUS can be useful in specific circumstances, these data suggest they present challenges in genetic testing for both patients and HCPs, with the potential to induce patient anxiety/stress. These results indicate support for providing the opportunity to opt-out of VUS return for MGP to patients, in the interest of patient autonomy, and to providers to reduce practice burden. The utility of VUS likely correlates with clinical context (e.g. cancer, rare disease, recessively versus dominantly inherited disorders etc). The clinical impact of VUS on patients/families warrants further investigation.

PrgmNr 3747 - Population genetic screening: What are the psychosocial impacts?

View session detail

Author Block: **N. Rao**[1], S. Coe[1], J. Huey[1], S. M. Fullerton[1], A. T. Chen[1], B. H. Shirts[2]; [1]Univ. of Washington, Seattle, WA, [2]Univ of Washington, Seattle, WA
Disclosure Block: **N. Rao:** None.
Population genetic screening has potential to identify individuals who carry a pathogenic variant putting them at risk for preventable hereditary cancers or hypercholesterolemia and who would otherwise be missed under current genetic testing guidelines. However, knowledge about the psychosocial impacts of receiving genetic results in a population screening setting is limited. We sought to understand these impacts among participants of the University of Washington Medical Centerâ□□s (UWMC) population genetic screening study.
This ongoing study invites adults who have received care at the UWMC to participate in targeted screening for hereditary cancers and familial hypercholesterolemia. Participants do not receive traditional genetic counseling prior to screening, but the study FAQ available upon joining details the type of information one might learn from the study as well as potential risks. Participants whose screening identifies no risk variants are notified of this information through an online report, while individuals whose results are positive for a risk variant are contacted by genetic counselors who describe the associated risks and emphasize the need for clinical follow-up. After receiving results, all participants are asked to complete an online survey about the psychosocial impacts of screening using the Feelings About GenomiC Testing Results (FACToR) instrument, which includes 4 subscales: negative emotions, positive emotions, uncertainty, and privacy. Participants can also respond to several open-ended survey questions about their response to results and related plans, and participants who screen positive may be selected for semi-structured interviews.
To date, 55% (N = 18) of participants who received positive screening results and 58% (N = 381) of participants who did not have completed the survey. On average, participants reported feeling some positive emotions and very little negative emotions, uncertainty or privacy concerns after receiving genetic results. Similar responses were seen for participants with positive results and those without. This may be because most individuals with concerns about adverse psychosocial impacts may have chosen not to participate in screening at all. In addition, sample size in the positive results group is limited and some members of this group were already aware of their genetic risk prior to screening. Overall, participants have not expressed any major concerns about receiving genetic results in this population screening setting.

PrgmNr 3748 - Preparing the next generation of healthcare providers for precision medicine in practice: A social justice and equity-oriented clinical genomics training program for pre-clerkship medical school students

View session detail

Author Block: **R. Rajagopalan**[1,2], T. J. Berninger[1,3], C. K. Rubanovich[1,4], C. S. Bloss[1,2]; [1]Ctr. for Empathy and Technology, T. Denny Sanford Inst. for Empathy and Compassion, Univ. of California, San Diego, La Jolla, CA, [2]Herbert Wertheim Sch. of Publ. Hlth.and Human Longevity Sci., Univ. of California, San Diego, La Jolla, CA, [3]Augustana-Sanford Genetic Counseling Graduate Program, Augustana Univ., Sioux Falls, SD, [4]Joint Doctoral Program in Clinical Psychology, San Diego State Univ./Univ. of California, San Diego, San Diego, CA
Disclosure Block: **R. Rajagopalan:** None.
As genomic medicine becomes increasingly widespread across clinical practice specialties, there remains an urgent yet largely unfulfilled need for delivering comprehensive and focused training in precision medicine tools, at all career stages in healthcare professions. At the same time, there have been calls to reform both genetics and medicine, recognizing that structural racism and implicit biases permeate clinical education, research, and practice, reinforcing persistent health disparities among historically disenfranchised, underserved, and marginalized communities. Few medical school training programs, however, currently offer educational content that addresses these needs, much less their intersection.

As part of efforts to bring issues of diversity, health equity, inclusion, and compassion to the fore of medical education in genomics, we describe the design, development, and preliminary evaluation of a novel training program called the Sanford Precision Health Scholars Immersive Learning Experience. This program aims to equip students at the earliest stages of their medical careers with the necessary skills for tackling challenges at the interface of precision medicine and anti-racist healthcare. Patient-driven case-based problem solving empowers students to build a toolkit for the cutting-edge of genomics practice, refracted through the principles of social justice and health equity. Interactive laboratory exercises with real-world genetic test reports anchor a set of learning objectives including: 1) awareness of the histories of racism and health disparities and relevance to contemporary calls for racial justice in genetics and medicine 2) familiarity with clinical genomics testing technologies and platforms 3) appreciation for the complexities of variant interpretation and classification 4) understanding the need for culturally appropriate, respectful, coordinated, and compassionate communication with diverse patients and families to promote psychosocial well-being as part of a clinical genetics care team, 5) knowledge of the benefits and limitations of patient-initiated DTC genomic testing for ancestry and health risks 6) recognition of the ethical dimensions of clinical decision-making around genetic testing, including the processes of informed consent, and the difficulties of interpreting test results especially among diverse communities that are underrepresented in genomic datasets. We discuss the implementation and evaluation of this program, prospects for stimulating student interest in clinical genomics careers, and strategies for scaling this program for replication at other institutions.

PrgmNr 3749 - Reframing translational research in rare diseases in the United States: A mosaic approach

View session detail

Author Block: **M. Halley**[1], H. S. Smith[2], A. Goldenberg[3], H. K. Tabor[4]; [1]Stanford Univ., Palo Alto, CA, [2]Baylor Coll. of Med., Houston, TX, [3]Case Western Reserve Univ, Cleveland, OH, [4]Stanford Univ. Sch. of Med., Stanford, CA

Disclosure Block: **M. Halley:** None.

The cumulative impact of rare diseases is immense, affecting an estimated 25 to 30 million Americans and contributing a tremendous physical, psychological and economic burden for patients, families and the healthcare system. Over 90 percent of rare diseases lack an approved therapy, and the pace of translating scientific discoveries into therapies is frustratingly slow. Recent advances in genome sequencing, molecular biology and machine learning suggest that real progress for the rare disease community could be on the horizon. However, our current approach to translational research in rare diseases in the United States lacks the integrated approach necessary to ensure that all patients and families in the rare disease community can benefit from these advances. In this paper, we lay out a new model for translational research in rare diseases, drawing on the concept of the mosaic, which emphasizes a multi-dimensional approach to identifying similarities within and across rare diseases. We argue that this new approach is essential not only to increase translational efficiency and maximize benefits across rare diseases, but also to move towards more equitable distribution of resources within the large and diverse rare disease community.

PrgmNr 3750 - Researcher Views on Multi-omics Return of Results in the Molecular Transducers of Physical Activity Consortium (MoTrPAC)

View session detail

Author Block: **C. Stanclift**[1], M. T. Wheeler[2], K. E. Ormond[3], J. N. Kohler[4], C. Reuter[5], M. E. Lindholm[6]; [1]Stanford Univ., Stanford, CA, [2]Stanford Univ, Palo Alto, CA, [3]Stanford Univ, Stanford, CA, [4]Palo Alto, CA, [5]San Francisco, CA, [6]Stanford Univ., Palo Alto, CA
Disclosure Block: **C. Stanclift:** None.
Introduction: There is growing consensus in favor of returning individual specific research results that are clinically actionable, valid, and reliable. However, deciding what and how research results should be returned remains a considerable challenge. Researchers are key stakeholders in return of results (ROR) decision-making and implementation. Multi-omics research contains medically relevant findings, therefore could be considered for return. We sought to understand researchers' views regarding the potential for multi-omics ROR.
Methods: We invited multi-omic researchers from the Molecular Transducers of Physical Activity Consortium (MoTrPAC) to participate. A brief recruitment survey was sent out via email to collect limited demographic information. A purposeful sampling of researchers participated in in-depth semi-structured interviews. To assess understanding of potential clinical utility for types of data collected and attitudes towards return of results for MoTrPAC clinical studies, an interview guide focusing on types of results generated in the study which could hypothetically be returned was devised from review of the literature and professional expertise of team members. The interviews were recorded, transcribed verbatim and co-coded, then discussed by the study team to identify thematic trends.
Results: A total of 16 researchers represented 11 sites and 6 consortium roles across MoTrPAC (co-investigator MD or MD, PhD 12.5%, co-investigator PhD 25%, recruitment/consent 18.8%, data production 12.5%, data analysis 25%, and exercise physiologist 12.5%). Many respondents supported hypothetical multi-omics ROR, citing participant rights to their own data and perception of minimal harm. However, ethical and logistical concerns around multi-omics ROR were raised including uncertain clinical validity, a lack of expertise to communicate results, and an unclear obligation regarding whether to return multi-omics results. Further, researchers called for more guidance from funding agencies and increased researcher education regarding ROR.
Conclusion: Overall, researchers expressed positive attitudes toward multi-omic ROR in principle, particularly if considered actionable. However, competing ethical considerations, logistical constraints, and need for more external guidance were raised as key concerns for implementation. Future studies should consider views and experiences of other relevant stakeholders, specifically clinical genomics professionals and study participants, regarding multi-omics ROR.

PrgmNr 3751 - Risk-reducing surgery in unaffected individuals receiving cancer genetic testing for hereditary breast and ovarian cancer in an integrated health care system

View session detail

Author Block: **B. Guo**[1], S. Knerr[2], K. F. Mittendorf[3], H. S. Feigelson[4], M. J. Gilmore[3], G. P. Jarvik[5], T. L. Kauffman[6], E. M. Keast[6], F. L. Lynch[6], K. R. Muessig[3], S. Okuyama[7], D. L. Veenstra[8], J. M. Zepp[3], K. A. B. Goddard[3], B. Devine[8]; [1]Dept. of Epidemiology, Univ. of Washington, Seattle, WA, [2]Dept. of Hlth.Services, Univ. of Washington, Seattle, WA, [3]Dept. of Translational and Applied Genomics, Ctr. for Hlth.Res., Kaiser Permanente Northwest, Portland, OR, [4]Inst. for Hlth.Res., Kaiser Permanente Colorado, Aurora, CO, [5]Dept. of Med., Div. of Med. Genetics, Univ. of Washington Med. Ctr., Seattle, WA, [6]Ctr. for Hlth.Res., Kaiser Permanente Northwest, Portland, OR, [7]Dept. of Med., Denver Hlth.and Hosp. Authority, Denver, CO, [8]Dept. of Pharmacy, Comparative Hlth.Outcomes Police and Economics Inst., Univ. of Washington, Seattle, WA

Disclosure Block: **B. Guo:** None.

Purpose: Germline genetic testing can identify high penetrance variants of hereditary cancers, enabling the early use of risk management options. Risk-reducing surgeries (RRS), such as bilateral mastectomy, bilateral salpingo-oophorectomy (BSO), and total hysterectomy decrease the risk of breast, ovarian, and endometrial cancer. However, prior studies on the uptake of RRS are often limited to those with pathogenic variant in well-studied genes within non-integrated healthcare systems. Therefore, we examined the uptake of RRS in unaffected individuals tested for multiple genes including those influencing breast, ovarian, and endometrial cancer risk between January 1, 2010 and December 31, 2018 in an integrated health care system. **Methods**: We conducted a retrospective cohort study of individuals aged **â¥**18 years without prior cancer history who received genetic testing for at least one hereditary cancer susceptibility genes within Kaiser Permanente Northwest. We described the uptake and timing of RRS by genetic test results (pathogenic/likely pathogenic (P/LP), variants of uncertain significance (VUS), and negative) using electronic health records and claims data. For those with P/LP, we further categorized surgery as recommended, recommended for consideration, or not recommended based on current National Comprehensive Cancer Network guidelines. **Results**: Of 1,020 individuals included, the mean age was 48 years (range: 18-â¥35 years, year of genetic testing, and increasing number of follow-up years. **Conclusion**: The uptake of RRS following hereditary cancer genetic testing was low among unaffected individuals in an integrated health care setting. Limitations of our data included no family history information and the inability to know whether surgery was informed by genetic test results or the reasons for not having/delaying surgery.

PrgmNr 3752 - SeqScreener: a cloud-based software application for analyzing CRISPR experiments with Sanger Sequencing data

View session detail

Author Block: **H. Fu**, Y. Chu; Thermo Fisher Bioscientific, South San Francisco, CA
Disclosure Block: **H. Fu:** Salary/Employment; Thermo Fisher Bioscientific.
Genome editing is poised to revolutionize the life sciences industry due to its precise changes at locus using CRISPR-Cas9 or other nucleases. SeqScreener was released to serve customers with data analysis of gene edit confirmation.Since the CRISPR repair process is not completely accurate, the pool of gene-edited cells needs rounds of screenings to become homogeneous. Pooled screen is to determine the relative fraction of desired edits. After cell expansion, a clonal screen finds clonally pure cells. As a critical step in screening protocols, Sanger sequencing is the most cost-effective measurement method with fast turnaround time to reveal the details around CRISPR cut site and the abundancy of each edited gene.Thermo Fisher Scientific had introduced True Design Genome Editor software for experiment designs. For data analysis, the SeqScreener Gene Edit Confirmation tool deconvolves the sequence trace of the mixture of wild-type and mutated sequences, generating results of gene editing efficiency, frame shift percentage, diversity of edits, and indel spectrum. It completed the Themo Fisher Scientific gene editing end-to-end workflow along with other gene editing products.Seqscreener has both pooled screening and clonal screening workflows. Its processing speed is 6.5 faster than competitors. It has clear visualization tools and the unique gene editing sample ranking system, which helps customers to select clones efficiently. SeqScreener also has capability to suggest related gene editing solutions to customers based on detected sample quality. The future plan is to provide â□□hands-freeâ□□ solutions such that sequencing data will be auto-uploaded directly from CE instruments and get analyzed within minutes without manual steps.SeqScreener simplifies data analysis and empowers the use of Sanger Sequencing as a genome editing confirmation tool.

PrgmNr 3753 - Splice-Switching Oligonucleotide-mediated Correction of a Deep Intronic Splice-variant in TIMMDC1 in Cells of Patients with Severe Early Onset Neurodegenerative Disorder

View session detail

Author Block: **J. Gecz**[1,2], M. A. Corbett[1], N. J. C. Smith[3,1], Z. Kikhtyak[1], A. Morimoto[1], S. Lee[4], J. G. Gleeson[5], E. A. Haan[6], R. Kumar[1]; [1]The Univ. of Adelaide, Adelaide, Australia, [2]South Australian Hlth.and Med. Res. Inst., Adelaide, Australia, [3]Women's and Children's Hlth.Network, Adelaide, Australia, [4]Univ. of California San Diego, San Diego, CA, [5]Univ California, La Jolla, CA, [6]Royal Adelaide Hosp., Adelaide, Australia
Disclosure Block: **J. Gecz:** Consultant/Consulting Fees/Other Remuneration; Royalty Free Consultant to Marinus Pharmaceuticals Ltd..
TIMMDC1 encodes the Translocase of Inner Mitochondrial Membrane Domain-Containing protein 1 (TIMMDC1) subunit of complex I of the electron transport chain responsible for ATP production. Variants causing loss of *TIMMDC1* expression have been implicated in a severe, progressive autosomal recessive disorder characterised by failure to thrive in the early postnatal period, poor feeding, hypotonia, peripheral neuropathy and drug resistant epilepsy. We have studied a consanguineous family with two affected children, now deceased, with clinical features overlapping *TIMMDC1* syndrome. Analysis of the genome sequencing data revealed previously reported pathogenic non-coding variant *TIMMDC1* c.597-1340A>G which is also present in gnomAD, (European non-Finnish population, 0.0002). Using RNAseq, on parentsâ and both patientsâ primary skin fibroblasts, we confirmed that this deep intronic *TIMMDC1* c.597-1340A>G variant enhances aberrant splicing, thereby inserting a poison exon that introduces a premature stop codon (p.Gly199_Thr200ins5*) into the *TIMMDC1* transcript that is eventually degraded via nonsense mediated mRNA decay. This leads to undetectable TIMMDC1 protein in patient fibroblasts and resulting in severely compromised mitochondrial complex I function. Carrier parents have reduced TIMMDC1 protein levels, but are unaffected. Subsequently, we have designed and applied two different splice-switching antisense oligonucleotides (SSO) to restore normal *TIMMDC1* mRNA processing and TIMMDC1 protein levels, in patients as well as in their carrier parentsâ fibroblasts. Seahorse real-time cell metabolic analysis of mitochondrial function in patient fibroblasts treated with SSOs also showed 100% restoration of complex I function. This research paves a way to the development of SSO-mediated therapy of this inevitably fatal, progressive neurological disorder, which is due, to likely ancestral, deep intronic variant with ~1/5000 allele frequency.

PrgmNr 3754 - An Embedded ELSI Framework for the Human Pangenome Reference Consortium

View session detail

Author Block: **A. B. Popejoy**[1,2], R. Cook-Deegan[3], N. A. Garrison[4], K. H. Miga[5], B. A. Koenig[6]; [1]Stanford Univ., Stanford, CA, [2]Univ. of California, Davis (UCD), Davis, CA, [3]Arizona State Univ., Accokeek, MD, [4]Univ. of California, Los Angeles (UCLA), Los Angeles, CA, [5]Univ. of California, Santa Cruz (UCSC) Genomics Inst., Santa Cruz, CA, [6]Univ. of California, San Francisco (UCSF), San Francisco, CA

Disclosure Block: **A.B. Popejoy:** None.

A comprehensive collection of diverse and highly accurate human reference genomes is critical to broaden our understanding of genetic variation and realize the promise of genomic medicine. The Human Pangenome Reference Consortium (HPRC) will use cutting-edge sequencing technologies to obtain high-quality phased genomes from a set of 350 diploid genomes, capturing and displaying much more human diversity than current reference sequences. This effort to diversify the human reference genome is arguably one of the most important initiatives in the field, as it has widespread implications for the practice of genomics research and the implementation of genetics in the clinic. While creating an open source, easily available global public data resource, the HPRC will simultaneously address critical ethical, legal, and social issues (ELSI) using an â□□embeddedâ□□ ELSI framework. The main objective of the HPRC-ELSI team is to identify, investigate, and ultimately offer consortium investigators advice about the most pressing issues they face, which must be addressed if the HPRC is to meet its goals. Our presentation will delineate the most immediate and fundamental questions facing the HPRC: 1) how should we define human populations for various purposes (i.e., recruitment, sampling, analysis, reporting) and determine what human â□□diversityâ□□ means in making a reference resource; 2) how can we assure truly informed consent for participation in a global resource; and 3) what constitutes robust data stewardship? By including insights from Indigenous data sovereignty scholars, we will present strategies to address the most pressing issues, concurrently explaining the full historical context of large-scale population genetics projects, such as the Human Genome Diversity Project. A full explication of the social context, including conflicting legal requirements in diverse locales or the political ramifications of misuse of genetic data for surveillance, will be paramount as the embedded ethics team offers guidanceâ□□in real timeâ□□to HPRC investigators. In the embedded model, with ELSI scholars participating in key meetings where decisions are made, investigators can engage these colleagues in discussions that deepen their understanding and appreciation of what is at stake as we seek to improve the human reference genome. As a result, the scientific goals of the HPRC will be substantively enriched through interdisciplinary scholarship that encourages collaboration, deliberation, and ultimately, shared decision-making.

PrgmNr 3755 - Determining Clinical Relevance of Genetic Variants in Newborn Screened Conditions

View session detail

Author Block: **A. M. Brower**; American Coll. of Med. Genetics and Genomics, Dakota Dunes, SD
Disclosure Block: **A.M. Brower:** None.
The goal of newborn screening (NBS) is to improve health outcomes by identifying and treating affected newborns. Each year over 22,000 newborns in the United States are identified with a genetic disease through NBS. NBS began in the 1960s with the discovery that newborns who received a screening test for phenylketonuria, using a blood spot on filter paper taken shortly after birth, benefited from early diagnosis and treatment. This discovery led to PKU screening pilots in several states and the eventual nationwide screening of all newborns using state-based public health laboratories. Over the past sixty years, this process of discovering the technologies to screen, diagnose, and treat has led to an increase in the number of screened conditions from 1 to 81. In addition, screening methodologies have advanced to include genomic sequencing, and treatment regimens have evolved to include gene-targeted therapies. The increasing use of genomics to screen, diagnose and treat conditions in newborns has heightened the need to understand the clinical relevance of genetic variants for newborn screened conditions. To aid in the interpretation of variants for NBS conditions, we created a computational method using longitudinal health information as part of the Newborn Screening Translational Research Network (NBSTRN). NBSTRN is an effort to develop data tools and resources to advance NBS related research and created the Longitudinal Pediatric Data Resource (LPDR). The LPDR enables the collection, storing, analysis, visualization, and sharing of genomic and phenotypic data. It is available for use by the research community and houses longitudinal clinical data on individuals with one of 118 rare genetic diseases. We performed a secondary analysis on data accumulated from a ten-year natural history study of 1904 individuals diagnosed with an inborn error of metabolism. We found 568 variants in 32 diseases and analyzed data on treatment and disease course. 28% (161/568) of the variants were not found in ClinVar and we were able to classify 86% (129/161) of these variants as pathogenic or likely pathogenic using the longitudinal health information. This work demonstrates that secondary analysis of longitudinal data collected as part of newborn screening finds unreported genetic variants. Further, the analysis of these databases provides the opportunity to reduce the number of variants of unknown significance, complicating interpretation and use of genomics to diagnose and treat individuals with genetic diseases.

PrgmNr 3756 - Experience from the 3D health pilot study at UCSF: how to implement and the utility of large scale genomic testing for healthy adult population

View session detail

Author Block: **J. Luevano**[1], S. Biswas[1], A. Rajkovic[2]; [1]Univ. of California, San Francisco, CA, [2]UCSF, San Francisco, CA
Disclosure Block: **J. Luevano:** None.
To better understand the predictive value of whole genome sequencing (WGS) in a healthy adult population, we launched a pilot sequencing study with an enrollment goal of 1,000 participants. We aim to establish institutional best practices to obtain, store, and communicate genomic information. In addition to understanding the utility of performing WGS for healthy adults, it is important to understand what type of genomic results participants want to receive, how to better engage clinicians (like primary care providers) and better engage participants of diverse backgrounds. In the pilot phase, we returned medically actionable variants of the 59 genes identified by the American College of Medical Genetics and an ancestry report. To date 494 adults [277 female (56.1%), 215 male (43.5%), 2 transgender (0.4%); Mean age 52.7 years; Ethnicity: 2 Alaska Native (0.4%), 82 Asian (16.6%), 13 Black or African American (2.6%), 18 Hispanic or Latino (3.6%), 1 Native Hawaiian or Other Pacific Islander (0.2%), 362 White/Caucasian (73.3%), 2 Mixed (0.4%), 7 Declined (1.4%), 11 Not available (2.2%)]. 494 patients have signed the consent form to participate in the study and 403 participantsâ samples have undergone whole genome sequencing, with 37 currently being processed. Of the 403 adults whose genomes were sequenced 11 (2.7%) were identified to have pathogenic (P) or likely pathogenic (LP) variants in the following genes: *BRCA1, KCNQ1, MYH7, APOB , CACNA1S, SCN5A, SDHAF2, RYR1 (2), TP53, BRCA2*. Participants were asked to complete a survey two weeks after receiving their results. Additionally, we have collaborated with the SPHERE (Special Populations and Health Equity in Research and Education) leaders to promote diversity among participants through community engagement. We have engaged with leaders among Black, Asian and Latinx communities, held town hall meetings and focus group studies to understand what the expectations, fears and obstacles are to participating in genomic research studies. The results from this pilot will help inform patient attitudes, develop a roadmap to scaling recruitment efforts and identify potential partners to help us further the effort.

PrgmNr 3757 - Functional annotation of human mitotic gene domains using CRISPR/Cas9 tiling screens

View session detail

Author Block: **P. J. Paddison**, L. Carter, S. Biggins, J. Herman; Fred Hutchinson Cancer Res. Ctr., Seattle, WA
Disclosure Block: **P.J. Paddison:** Major Stockholder/Ownership Interest; Presage Biosciences. Royalty(ies)/Honoraria; RNAi/shRNA Royalties. Consultant/Consulting Fees/Other Remuneration; Cellinta. Receipt of Intellectual Property Rights/Patent Holder; shRNA patents.
A critical knowledge gap for the human genome has arisen from our inability to resolve important functional domains and motifs within protein coding genes at the large scale. Historically, large scale annotation of protein domains and motifs relied on homology based-inference by searching against the current 5494 conserved protein family (Pfam) domains documented in the human genome (e.g., methyltransferase-like domain). This approach is ineffective for the ~45% of the proteome that is devoid of Pfam domains, and still requires validation for the remaining genes. Closing this knowledge gap is critical for both basic and disease-focused biomedical research, where years, if not decades, can be spent dissecting gene functions. Here, we develop a method that leverages CRISPR-Cas9 induced mutations to comprehensively identify key functional regions within protein sequences of the human genome that are required for cellular outgrowth. Our analysis of 48 human kinetochore genes revealed hundreds of regions required for cell proliferation, corresponding to known and uncharacterized protein domains. We biologically validated 15 of these regions, including amino acids 387-402 of Mad1, which when deleted compromise Mad1 kinetochore localization and chromosome segregation fidelity. Altogether, we demonstrate that CRISPR-Cas9-based tiling mutagenesis enables *de novo* identification of key functional domains in protein-coding genes, which elucidates separation of function mutants and allows functional annotation across the human proteome.

PrgmNr 3758 - IEMbase: An online knowledge base and mini-expert tool for the diagnosis of inborn errors of metabolism

View session detail

Author Block: **B. Hewitson**[1], N. Blau[2], W. W. Wasserman[3], C. D. van Karnebeek[4], K. Ã□unap[5], C. R. Ferreira[6]; [1]Univ. of British Columbia, Vancouver, BC, Canada, [2]Univ. Children's Hosp., Zurich, ZH, Switzerland, [3]Univ British Columbia, Vancouver, BC, Canada, [4]Amsterdam UMC, Amsterdam, Netherlands, [5]Univ. of Tartu, Tartu, Estonia, [6]Bolingbrook, IL
Disclosure Block: **B. Hewitson:** None.
Inborn errors of metabolism (IEMs) are inherited disorders that disrupt metabolic pathways as a result of defects in the genes that code for enzymes involved in facilitating the process. The bodyâ□□s inability to successfully synthesize metabolites from substrates can lead to an accumulation of toxic substrates or a deficiency of essential metabolites. Thus, a delay in the diagnosis of IEM disorders can lead to severe consequences for patients such as irreversible organ damage or death. Therapy is available to improve patient outcomes for a number of these rare inherited disorders; however, an accurate and timely diagnosis is required for effective treatment. This is often difficult to achieve due to a large knowledge gap between IEM specialists and the clinicians involved in early diagnoses, as well as nonspecificity of symptoms for IEM disorders. As such, we present the latest version of IEMbase, an online tool aimed at bringing the knowledge of specialized experts to professionals involved in diagnoses. IEMbase consists of an expert-curated knowledge base and mini-expert diagnosis support system, which accepts a list of clinical and/or biochemical phenotypes from the user and returns a list of matching IEM diagnoses. The matched diagnoses are generated using a two-step algorithm to compare the input phenotypic profile with every IEM profile in the IEMbase database. The first step assesses biochemical phenotypes by calculating the cosine similarity of term frequency-inverse document frequency (TFIDF) vectors for the input profile and each IEM profile in the database. The resulting cosine similarity scores are used to rank IEM profiles in the database to determine the disorders that are to be used in the second step of the algorithm. The semantic similarity between the input clinical phenotype and IEM disorder profiles in the database are calculated in the second step of the algorithm, relying on the Human Phenotype Ontology (HPO) as the standard medical vocabulary. The disorders resulting from the mini-expert system can be used to generate differential diagnosis tables, biochemical test panels, and targeted gene panels for users to pursue further investigations for diagnosis. Approximately 400 monthly users access the information in IEMbase, currently consisting of 3430 clinical/biochemical profiles and 8465 disorder/phenotype profiles created from 1540 IEM disorders.

PrgmNr 3759 - Rarediseasegenes.com/phex: A comprehensive locus specific database of *PHEX* gene variants associated with X-linked hypophosphatemia vastly increases the number of known variants

View session detail

Author Block: **S. Daugherty**[1], S. Sarafrazi[1], N. Miller[1], P. Boada[1], T. O. Carpenter[2], L. Chunn[3], M. J. Econs[4], S. Eisenbeis[1], E. A. Imel[4], B. Johnson[5], M. J. Kiel[3], S. Krolczyk[1], P. Ramesan[1], R. M. Truty[5], Y. Sabbagh[6]; [1]Ultragenyx Pharmaceutical, Novato, CA, [2]Yale Univ. Sch. of Med., New Haven, CT, [3]Genomenon, Ann Arbor, MI, [4]Indiana Univ. Sch. of Med., Indianapolis, IN, [5]Invitae Corp., San Francisco, CA, [6]Inozyme Pharma, Boston, MA
Disclosure Block: **S. Daugherty:** Salary/Employment; Ultragenyx Pharmaceutical.
X-linked Hypophosphatemia (XLH), a dominantly inherited disorder caused by loss of function mutations in the *PHEX* gene, is characterized by rickets in childhood followed by progressive skeletal disease. Short stature and deformity often ensue, with lifelong impairment of mobility and quality of life. Early and accurate diagnosis based on clinical, biochemical, and genetic features is complicated by the rarity of the disease, phenotypic variability, and similarities to other forms of congenital and sporadic hypophosphatemia. Advances in sequencing methodologies have led to detection of increasing numbers of *PHEX* variants. A comprehensive summary of *PHEX* disease-associated variants is necessary for accurate and timely interpretation of genetic testing results. We developed a comprehensive locus-specific database (LSDB) for *PHEX* to collect and disseminate information about *PHEX* variants to the scientific community. The database was developed by integrating variants from four different sources: **1.** an archived McGill University *PHEX* LSDB (Sabbagh et al., *Hum Mol Genet.,* 2001.); **2.** a comprehensive, systematic review of all published medical literature; **3.** a sponsored hypophosphatemia gene panel testing program (Rush et al., manuscript in preparation); and **4.** Ultragenyx clinical programs. Variants were annotated with detailed clinical and biochemical phenotypes extracted from literature reports and clinical testing submission forms. As of 28 APR 2021, the *PHEX* LSDB reports 870 unique *PHEX* variants, forming 869 alleles. Single nucleotide variants (n= 321, 36.90%) represent the most common type of unique variant in the database followed by small deletions, duplications, or insertions (n= 282, 32.41%), splicing variants (n= 191, 21.95%), and copy number variants (n= 75, 8.02%). This represents a >200% increase in the number of *PHEX* variants documented in the original archived PHEX LSDB (last updated 2APR2017). Most of the variants have an ACMG call or predicted call of pathogenic or likely pathogenic (n=689, 79.20%) and 167 (19.20%) were categorized as variants of uncertain significance. The comprehensive searchable database of *PHEX* variations is freely available (www.rarediseasegenes.com/phex) and provides an opportunity to add new variants, which will increase the diagnostic yield from genetic testing and be an important tool for XLH research.

PrgmNr 3760 - Re-architecture of ClinGenâ□□s Variant Curation Interface (VCI) to Support Future Enhancements Aimed at Scaling Variant Classification

View session detail

Author Block: **C. Preston**[1], M. Wright[1], L. Madhavrao[1], G. Cheung[1], M. E. Mandel[1], H. Tong[1], B. Wulf[1], S. Harrison[2], J. Goldstein[3], X. Luo[4], M. DiStefano[5], D. I. Ritter[4], K. Riehle[6], N. Shah[4], A. Milosavljevic[4], J. S. Berg[7], H. L. Rehm[8], S. E. Plon[9], H. A. Costa[10], C. D. Bustamante[1], T. Klein[1], Clinical Genome (ClinGen); [1]Stanford Univ. Sch. of Med., Stanford, CA, [2]Broad Inst., Cambridge, MA, [3]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [4]Baylor Coll. of Med., Houston, TX, [5]Geisinger, Danville, PA, [6]Baylor Coll. of Med., houston, TX, [7]UNC Chapel Hill, Chapel Hill, NC, [8]Massachusetts Gen. Hosp., Boston, MA, [9]Baylor Coll. Med., Houston, TX, [10]Stanford Univ, Stanford, CA

Disclosure Block: **C. Preston:** None.

The ClinGen Variant Curation Interface (VCI) is a global open-source variant classification platform for supporting the application of evidence criteria and classification of variants based on the ACMG/AMP sequence variant classification guidelines. The VCI is among a suite of tools developed by the NIH-funded Clinical Genome Resource (ClinGen) Consortium, and supports an FDA-recognized human variant curation process. To facilitate evidence-based improvements in human variant classification, the VCI is publicly available to the genomics community. ClinGen is expanding to involve more curators and an increased scale of activities as we include more genetic variants across a greater number of genes. Here we present interface design decisions made during a recent VCI platform re-architecture that will support future enhancements and enable scalable curation workflows. The new interface system involves moving from a server-based cloud architecture to a serverless architecture. Evidence codes that are amenable to automated application (e.g. allele frequency, predicted loss of function, in silico prediction) will be supported in future versions of the VCI as well as enabling gene-specific specifications and thresholds for applying evidence criteria. These developments support the ClinGen mission of curating the continually-expanding space of the clinical human genome.

PrgmNr 3761 - Significant Sparse Polygenic Risk Scores in 428 traits in UK Biobank

View session detail

Author Block: **Y. Tanigawa**, J. Qian, R. Li, G. R. Venkataraman, J. M. Justesen, R. Tibshirani, T. J. Hastie, M. Rivas; Stanford Univ., Stanford, CA
Disclosure Block: **Y. Tanigawa:** Consultant/Consulting Fees/Other Remuneration; Cipherome, Inc.
Polygenic risk scores (PRSs) have been proposed for disease risk prediction for some traits, but their predictive ability across human traits remains unclear. Most approaches use many variants from GWAS analyses, limiting their biological interpretability. Here, we systematically characterize an atlas of sparse PRSs in more than 1,500 traits (> 450 disease outcomes) in the UK Biobank with penalized multivariate regression using the R snpnet package (batch screening iterative Lasso, BASIL), which simultaneously performs variable selection and effect-size estimation using individual-level data. Using a held-out test set, we systematically compare the predictive ability of the PRSs to baseline models generated with covariates (age, sex, and genotype PCs) and find significant (p -5) improvements in predictive performance in 428 traits. We assess the trans-ethnic predictive performance within the UK Biobank and find limited transferability across ancestry groups for most traits. The results of our systematic evaluation of sparse PRS and the coefficients of each model are publicly available on Global Biobank Engine (https://biobankengine.stanford.edu/prs).

PrgmNr 3762 - Specifications of the ACMG/AMP variant curation guidelines for *MYOC*: Recommendations from the ClinGen Glaucoma Expert Panel

View session detail

Author Block: **E. Souzeau**[1], K. P. Burdon[2], J. E. Craig[1], P. Graham[2,3], J. Hadler[1,4], A. W. Hewitt[2], J. D. Hulleman[5], D. A. Mackey[2,3], E. Nakahara[5], F. Pasutto[6], O. M. Siggs[1,7], T. L. Young[8], A. Dubowsky[4]; [1]Flinders Univ., Adelaide, Australia, [2]Univ. of Tasmania, Hobart, Australia, [3]Univ. of Western Australia, Perth, Australia, [4]SA Pathology, Adelaide, Australia, [5]Univ. of Texas Southwestern Med. Ctr., Dallas, TX, [6]Univ. Erlangen-Nuremberg, Erlangen, Germany, [7]Garvan Inst. of Med. Res., Sydney, Australia, [8]Univ. of Wisconsin-Madison, Madison, WI

Disclosure Block: **E. Souzeau:** None.

Background: Standardised variant curation criteria are essential for accurate interpretation of genetic results and clinical care of patients. The variant curation guidelines developed by ACMG/AMP in 2015 are widely used by the genetic community but are not gene specific. The Clinical Genome Resource (ClinGen) Variant Curation Expert Panels (VCEP) are tasked with developing gene-specific variant curation guidelines. *Aim & Methods:* The Glaucoma VCEP was created in 2019 and assembled clinicians, researchers and laboratory scientists that aimed to develop and pilot rule specifications of the ACMG/AMP variant curation guidelines for *MYOC* variants, the most common cause of Mendelian glaucoma. *Results:* Among the 28 ACMG/AMP criteria, the Glaucoma VCEP adapted 15 rules to *MYOC* (including 10 strength specifications), while 13 rules were deemed not applicable. Key specifications included calculations of minor allele frequency thresholds, developing an approach to counting multiple independent probands and segregations, and reviewed functional assays that reported on the solubility and secretion of mutant proteins. The specified rules were piloted on 82 variants and led to a change in classification in 14/37 (38%) of those that were classified in ClinVar. Functional evidence led to the reclassification of 19 (44%) of VUS (including 13 from VUS to likely pathogenic and 6 from VUS to likely benign). *Conclusion:* The standardised variant curation guidelines for *MYOC* from the Glaucoma VCEP improved variant classification. *MYOC* variant curation and classifications will be submitted to ClinVar with expert level status.

PrgmNr 3763 - The Atlas of Variant Effects (AVE) Alliance:understanding genetic variation at nucleotide resolution

View session detail

Author Block: **L. Starita**[1], M. E. Hurles[2], D. J. Adams[3], A. L. Gloyn[4], W. C. Hahn[5], D. Marks[6], J. T. Neal[7], F. P. Roth[8], A. F. Rubin[9], D. M. Fowler[1]; [1]Univ of Washington, Seattle, WA, [2]Wellcome Sanger Inst., Cambridge, United Kingdom, [3]Wellcome Sanger Inst., Hinxton, United Kingdom, [4]Stanford Univ., Stanford, CA, [5]Dana Farber Cancer Inst., Boston, MA, [6]Harvard Med. Sch., Boston, MA, [7]Broad Inst., Cambride, MA, [8]Univ. of Toronto, Toronto, ON, [9]Walter and Eliza Hall Inst., Melbourne, Australia
Disclosure Block: **L. Starita:** None.
It has become nearly routine to sequence human exomes and genomes. However, functional impacts are characterized for The Atlas of Variant Effects (AVE) Alliance is a new initiative aimed at comprehensively measuring the biological effect of genetic variation in human and pathogen genomes, to advance clinical diagnosis and disease management, drug development and fundamental mechanistic understanding. We share a long-term vision to generate and apply an Atlas of thousands of variant effect maps for both proteins and non-coding elements.
The AVE Alliance will support and expand the community generating variant effect maps, by coordinating efforts and establishing standards for generating, managing, sharing and using variant effect map data responsibly and with maximum impact. Please see varianteffect.org to learn more and consider joining AVEâ□□s current ranks of 180 members internationally.

PrgmNr 3764 - Virtual Registry of 3,500 Leigh Syndrome and Primary Mitochondrial Disease Cases Constructed through Expert Curation and Semi-automated Literature Mining at MSeqDR

View session detail

Author Block: **L. Shen**[1], M. T. Lott[2], E. M. McCormick[3], C. C. Muraresku[3], K. N. Keller[4], G. ClinGen_Mitochondrial_Diseases_GCEP[1], D. C. Wallace[5], Z. Zolkipli-Cunningham[6], S. Rahman[7], M. J. Falk[8], X. Gai[1]; [1]Children's Hosp. Los Angeles, Los Angeles, CA, [2]Ctr. for Mitochondrial and Epigenomic Med., Div. of Human Genetics, Dept. of Pediatrics, Childrenâ□□s Hosp. of Philadelphia, Philadelphia, PA, [3]Dept. Mitochondrial Med. Frontier Program, Childrenâ□□s Hosp. of Philadelphia, Philadelphia, PA, [4]Children s Hosp. of Philadelphia, Philadelphia, PA, [5]CHOP - Univ Pennsylvania, Philadelphia, PA, [6]CHOP, Colket Translational Res. Building, 6th floor, Philadelphia, PA, [7]UCL Great Ormond Street Inst. of Child Hlth., London, United Kingdom, [8]Ctr. for Mitochondrial and Epigenomic Med., Div. of Human Genetics, Dept. of Pediatrics, 3Dept. Mitochondrial Med. Frontier Program, Dept. of Pediatrics, Childrenâ□□s Hosp. of Philadelphia, Philadelphia, PA

Disclosure Block: **L. Shen:** None.

Primary mitochondrial disease (PMD) has a prevalence of at least 1 in 4,300 people. The most common pediatric manifestation of PMD is Leigh Syndrome (LS) with a prevalence of 1 in 34,000 people and more than 100 associated gene causes now recognized. PMD community registry creation has required intensive collaborative effort, with 1,555 participants enrolled in the North American Mitochondrial Disease Consortium (NAMDC) Registry as of 2020 (https://ng.neurology.org/content/6/2/e402).

To enhance knowledge of PMD, we built at MSeqDR a PMD Virtual Registry using both semi-automated literature mining and expert review. Specifically, we developed a semi-automated pseudo-case curation platform, hosted at MSeqDR.org, to aid in the capture, extraction, and standardization of case-level data from publications through semantic matching followed by manual review. In parallel, the NIH funded ClinGen Mitochondrial Diseases Gene Curation Expert Panel (HD-093483, https://www.clinicalgenome.org/affiliation/40027/, https://reporter.nih.gov/project-details/9411950) manually curated ~350 deeply-phenotyped cases associated with 113 nuclear and mtDNA genes from published literature to evaluate the causative gene associations for Leigh Syndrome spectrum (LSS) disorder. Heterogeneous data were mapped to over 100 â□□standardâ□□ clinical, biochemical, and genetic feature terms. Case level data was fully atomized, and divided into over 110,000 atomized feature records. Phenotypes and diseases were mapped to Human Phenotype Ontology (HPO) and OMIM terms.

The Web front-end includes a PMD Virtual Registry Case Browser, which supports querying using single or composite filters constructed from atomized and standardized keywords such as OMIM disease, (causative) genes and variants, phenotypes and HPO terms, inheritance mode, zygosity, consanguinity, sex, age at onset, age at death, ethnicity, PubMed ID and title, and date ranges. Matching cases are hyper-linked to single case report pages, which presents all data elements of a case in both standardized and original terms.

As of June 2021, the MSeqDR PMD Virtual Registry contains ~1,050 LS/LLS cases (>700 from semi-automatic literature mining plus 350 from U24 project expert curation) plus 2,450 other PMD cases. This collection of 3,500 de-identified PMD pseudo-cases exceeds most existing ad-hoc LS/LSS or PMD registries. Full curation of the pseudo cases and in-depth stratification and correlation analysis are ongoing. The PMD Virtual Registry is hosted on the United Mitochondrial Disease Foundation (UMDF) supported MSeqDR platform (https://mseqdr.org/virtualregistry.php).

PrgmNr 3765 - A rare case of Bardet Biedl syndrome presented as acute onset of heart failure and dilated cardiomyopathy

View session detail

Author Block: **H. Wang**[1], A. Stacy[1], A. Munoz[1], J-A. Gold[1,2]; [1]Dept. of Pediatrics, Loma Linda Univ., Sch. Of Med., Loma Linda, CA, [2]Dept. of Pediatrics, Div. of Genetics and Genomics, Univ. of California Irvine, Irvine, CA
Disclosure Block: **H. Wang:** None.
INTRODUCTION: Bardet-Biedl syndrome (BBS) is a rare genetically heterogeneous, autosomal recessive inherited disorder with wide variability in expression. Characteristic features of BBS include retinitis pigmentosa, post axial polydactyly, central obesity, and learning disability. Other manifestations include diabetes mellitus, heart disease, hepatic fibrosis, renal disease and neurological manifestations. BBS presenting as acute onset of heart failure and dilated cardiomyopathy is very rare. **CASE DESCRIPTION:** We present the case of a 11-year-old female, with history of intermittent asthma, obesity, developmental delay (~2nd-3rd grade level), She was admitted for new-onset acute systolic heart failure. She was hypoxic and CXR demonstrated a large right pleural effusion and cardiomegaly. BNP elevated to 3959. Severe LV dysfunction with LV 24%, dilated cardiomyopathy with biventricular dysfunction on her TTE. Right pleural chest tube was placed and found to have chylous effusion. She was intubated and ECMO was initiated. Cardiac biopsy showed hypertrophy and liver biopsy indicated bridging fibrosis consistent with cirrhosis. The previous medical records from another state showed three variants in BBS1 gene: 1). c.1169T>G, (p.Met390Arg), heterozygous, recessive, associated; 2). c.433-2A>C, splice-site mutation, heterozygous, recessive, probably associated. 3). c.7A>G, 3'UTR mutation, heterozygous, not associated. Her brother was tested first he has the same variants. She was diagnosed with BBS. On physical exam, she is tall in 95% percentile with BMI 36, no obvious dysmorphic features were appreciated. The patient subsequently underwent cardiac transplant and is currently in the early postoperative recovery period. **DISCUSSION/CONCLUSION**: Bardet-Biedl syndrome is a rare autosomal recessive multisystem disorder caused by defects in genes encoding for proteins that localize to the primary cilium/basal body complex. A review of the literature did not find reports of BBS with cardiac transplant. More than 20 disease-causing genes have been identified to date. Congenital and acquired heart disease have been reported but cardiac monitoring is not required in surveillance plan. Echocardiographic examination should be included in the clinical evaluation and follow-up of patients with BBS.

PrgmNr 3766 - Further delineating Baraitser Winter syndrome through ACTB and ACTG1 phenotypic characterization

View session detail

Author Block: **L. Forero**[1], A. Lee[2], L. M. Bird[3]; [1]Univ. of California San Diego, San Diego, CA, [2]Univ. of California San Diego, San Diego, CA, [3]Rady Children's Hosp., San Diego, CA
Disclosure Block: **L. Forero:** None.
Baraitser-Winter syndrome (BWS) is characterized by multiple congenital anomalies, including facial dysmorphism, developmental delay, CNS abnormalities, and short stature. BWS is an autosomal dominant disorder caused by mutations in two genes, *ACTB* and *ACTG1*. We performed a comprehensive review of the literature searching for cases of BWS with molecular confirmation and included two patients seen through our institution. This resulted in a cohort of 76 patients, for which we created a database of 31 HPO terms specifying CNS, eye, facial, cardiovascular, hematologic, and growth/limb anomalies. Our goal was to study genotype-phenotype correlation and to assess the severity associated with genotype. Of the 76 patients in our cohort, 24 carried a mutation in *ACTB* and 52 in *ACTG1*. Two by two tables for every HPO term were constructed, and Fisher exact test was used to analyze the data. P ACTG1 group and only 38% (20/25) of patients who carried mutations in *ACTB*. Lissencephaly was reported in 67% (16/24) of patients with mutations in *ACTG1*compared to 31% (16/52) of *ACTB* patients. For *ACTG1* patients, hydrocephalus was present in 21% (5/24), while it was only documented in 4 % (2/52) of patients from the *ACTB* group. These findings are concordant with prior studies reporting a more severe CNS phenotype for patients with mutations in *ACTG1*. Of note, there was no statistically significant difference in intellectual disability/ developmental delay (DD) between the two genes. ID/DD was present in 75% (18/24) of *ACTG1* patients and 60% of patients in the *ACTB* group. We were not able to determine the severity of ID/developmental. To our knowledge, this is one of the largest BWS cohorts looking for genotype/phenotype correlation. It appears that overall, there is significant overlap between the two genotypes except for the mentioned CNS and eye anomalies. We will discuss the reported variant types/functional consequences and their correlation with the severity of phenotype during our presentation.

PrgmNr 3767 - Lessons learned from the first 170 trios from the Genomic Autopsy Study

View session detail

Author Block: **C. Barnett**[1], T. Ha[2], P. Arts[3], A. B. Byrne[4], M. Babic[2], L. Pais[5], A. Schreiber[6], G. Genomic Autopsy Study Research Network[1], K. Kassahn[2], T. Hardy[7], M. Jackson[1], H. S. Scott[7]; [1]Women's and Children's Hosp., North Adelaide, Australia, [2]Genetics and Molecular Pathology Res. Lab., Ctr. for Cancer Biology, An alliance between SA Pathology and the Univ. of South Australia, Adelaide, Australia, North Adelaide, Australia, [3]Ctr. for Cancer Biology, Adelaide, Australia, [4]Henley Beach, Australia, [5]Ctr. for Mendelian Genomics Broad Inst., Boston, MA, [6]ACRF Cancer Genomics Facility, Ctr. for Cancer Biology, An alliance between SA Pathology and the Univ. of South Australia, Adelaide, Australia, North Adelaide, Australia, [7]SA Pathology, Adelaide, Australia

Disclosure Block: **C. Barnett:** None.

Background: The cause of pregnancy loss and perinatal death remains unexplained in at least 25% of cases, despite a high perinatal autopsy rate in Australia. The Genomic autopsy study is a national collaborative study aimed at determining the cause. Aim: To apply WES/WGS to identify genetic causes of fetal/newborn abnormalities that result in termination of pregnancy, death in utero or in the newborn period, in view to providing families with answers regarding cause and likelihood of recurrence. Methods: WES/WGS is being performed on parent-fetus trios/quads following standard autopsy and non-informative microarray. High priority cases are consanguineous families, fetuses with multiple malformations, and unexplained fetal/newborn death. Statistical, bioinformatic and experimental laboratory techniques are used to confirm causality of variants. Results: 170 prospective trios (150) or quads (20) have been recruited and sequenced. 89/170 (52%) are either solved or have a strong candidate gene identified. 42/170 (25%) of cases have been clearly solved by identification of a pathogenic mutation in a known gene. An additional 27/170 (16%) have a candidate variant in a known gene that expands phenotype. In 20/170 (12%) of cases, a mutation was identified in a gene not yet linked to human disease. Unexplained stillbirth cases remain the most difficult to solve. Discussion: Our results provide justification for genomic investigation of pregnancy loss and perinatal death, particularly when congenital abnormalities are present. Establishing a clear diagnosis on clinical grounds alone is often challenging. Late stillbirth remains largely unexplained by genomics. Novel techniques for genomic investigation of unexplained stillbirth need further consideration, including genomic investigation of the placenta.

PrgmNr 3768 - Parental Mosaicism in 'De Novo' Tuberous Sclerosis Complex

View session detail

Author Block: **z. ye**[1], S. Lin[2], X. Zhao[2], Y. Yao[2], L. Li[2], L. Chen[2], J. Duan[2], D. Cao[2], Z. Hu[2], S. Berkovic[1], I. Scheffer[1], J. Liao[2], M. Hildebrand[1]; [1]Univ. of Melbourne, Mebourne, Australia, [2]Shenzhen Children's Hosp., Shenzhen, China
Disclosure Block: **Z. ye:** None.
Rationale: Tuberous sclerosis complex (TSC) is a genetic disorder associated with neurological, renal, dermatological and other anomalies. Over 85% of reported patients have germline mutations in *TSC1* or *TSC2*, with a higher yield of pathogenic variants in cohorts with more severe, earlier onset disease. Some patients appear to have â□□de novoâ□□ variants on routine testing of parental blood-derived DNA, yet one parent may have clinical features of TSC or multiple affected children carrying the same germline mutation, indicating the likelihood of low-level parental mosaicism. We aimed to identify parental mosaicism in 7 families with TSC and features suggesting underlying parental mosaicism.
Methods: Seven families had known pathogenic germline mutations previously identified in the probands on clinical genetic testing. All parents were negative for their childâ□□s variant on segregation analysis via clinical Sanger sequencing. Blood, saliva, buccal and urine DNA were obtained from all parents. Droplet digital PCR (ddPCR) or deep targeted amplicon sequencing (5,000x) were used to detect and quantitate the variants in parental tissues.
Results: We found parental mosaicism in blood-derived DNA in five of 7 families at low frequency ranging from 0.1-8.8%, all well below the threshold of detection by Sanger sequencing. In 3 parents with mosaic variants, we determined variant allele frequency (VAF) in different tissues. One parent had mosaicism at a similar level across four different tissues (7.26-9.11% VAF); while in the other two parents, mosaicism was variable (0.9-3.12% VAF across three tissues and 0.39-2.2% VAF across four tissues, respectively). Of the two families without mosaicism detected, one had two affected children carrying the same germline mutation so one parent must have gonadal mosaicism that we could not detect in peripheral tissues. In the second negative family, the mother has unilateral renal hamartoma without any other clinical features of TSC; so, her hamartoma may not be due to TSC.
Conclusions: Our findings confirm that low-level parental mosaicism missed on routine clinical testing can be detected with high levels of coverage. This finding has critical implications for reproductive counseling.

PrgmNr 3769 - *De novo* heterozygous *POLR2A* variant associates with autism spectrum disorder with epilepsy, hypotonia, strabismus and self-injurious behaviours

View session detail

Author Block: **D. Evans**[1], Y. Qiao[2], S. Martel[2], K. Calli[2], S. Jones[2,3], S. Scherer[4], S. M. Lewis[5]; [1]Dept. of Family Practice, Univ. of British Columbia, Vancouver, BC, Canada, [2]Med. Genetics, Univ. of British Columbia, Vancouver, BC, Canada, [3]Canadaâ□□s Michael Smith Genome Sci. Ctr., Vancouver BC, BC, Canada, [4]The Ctr. for Applied Genomics and McLaughlin Ctr., Hosp. for Sick Children and Univ. of Toronto, Toronto, ON, Canada, [5]BC Children's & Women's Hlth.Ctr, The Univ. of British Columbia, Vancouver, BC, Canada

Disclosure Block: **D. Evans:** None.

Autism spectrum disorder (ASD) describes a complex and heterogenous group of neurodevelopmental disorders. Whole genome sequencing studies are beginning to elucidate the multifactorial etiology of ASD. Studies employing pathway analysis for genetic variants in ASD have suggested that dysregulation and dysfunction of transcriptional pathways are linked to neurodevelopmental disorders. *POLR2A* encodes the RBP1 subunit of the RNA Polymerase II Complex (pol II). This fundamentally important enzyme transcribes all ~21,000 protein coding genes. Recent studies demonstrate a putative role for *POLR2A* in neurodevelopmental disorders. In particular, *de novo* *POLR2A* variants were recently implicated in a new neurodevelopmental phenotype called â□□Neurodevelopmental Disorder with Hypotonia and Variable Intellectual and Behavioral Abnormalitiesâ□□ (NEDHIB) (OMIM: 618603). Patients described with NEDHIB are characterized by early-onset hypotonia, delayed walking, poor speech, variable intellectual development impairment, and behavioral abnormalities with a high degree of variability in severity. The disorder segregates in an autosomal dominant manner. In this study, we describe and characterize a novel *de novo* deleterious *POLR2A* variant (c.1367T>C, p. Val456Ala) identified in an 8-year-old girl with ASD. The variant was detected using proband-parent trio whole genome sequencing. In this simplex case, the *POLR2A* p.Val456Ala variant manifests with high functioning ASD, central hypotonia, seizure, macrocephaly, hyperopia, strabismus, and self-injurious behaviour with no obvious dysmorphism aside from subtle right sided facial prominence. She presented with speech delay with flat intonation, and heightened sensitivity to touch, sound, and texture. This study provides further evidence for the role of *POLR2A* in neurodevelopmental disorders, while particularly emphasizing the role for this gene in ASD. Mounting evidence now suggests that disruption of *POLR2A* manifests as a spectrum of multi-system developmental disorders.

PrgmNr 3770 - A CRISPR screen for modifiers of the rare disease DPAGT1-CDG (CDG-Ij)

View session detail

Author Block: **H. M. Dalton**[1], R. Viswanatha[2], R. Brathwaite Jr.[2], S. E. Mohr[2], N. Perrimon[2,3], C. Y. Chow[1]; [1]Univ. of Utah, Salt Lake City, UT, [2]Harvard Med. Sch., Boston, MA, [3]Howard Hughes Med. Inst., Boston, MA

Disclosure Block: **H.M. Dalton:** None.

Partial loss-of-function mutations in glycosylation pathways underlie a set of rare diseases called Congenital Disorders of Glycosylation (CDGs). Glycosylation is a broad category of sugar modifications on proteins and lipids, with functions ranging from complex post-translational modifications through the endomembrane system (e.g. N-glycosylation) to single sugar additions involved in cell signaling (e.g. O-GlcNAc-ylation). CDGs have a range of symptoms, but commonly include severe epilepsy, developmental delay, and disability. CDG Type Ij is caused by loss-of-function mutations in *DPAGT1* - the first step in N-glycosylation. Our goal is to better understand the pathways connected to *DPAGT1* loss, as well as glycosylation disorders as a whole, to develop potential treatment options.

We performed a CRISPR knockout screen using the drug tunicamycin (Tun), a potent inhibitor of DPAGT1 function, on *Drosophila* S2R+ cells. Loss of *DPAGT1* and Tun treatment impair N-glycosylation and cause massive protein misfolding, leading to reduction of cell surface glycoproteins and endoplasmic reticulum (ER) stress. Because many CDGs are caused by mutations in the N-glycosylation pathway, characterizing the genes that affect this pathway will help us better understand these disorders. Using a pooled format, we introduced a whole genome guide RNA library into the S2R+ cells expressing constitutive Cas9. Pooled cell populations were grown for 10 generations under selection with either vehicle or Tun. Final populations were examined for total guide RNA abundance to determine candidate genes causing resistance or sensitivity to *DPAGT1* inhibition.

We validated candidate genes by Concanavalin A (ConA) to assay cell surface glycoproteins in the S2R+ cells, and by RNAi knockdown via an *in vivo* ER stress *Drosophila* model. The ConA assay demonstrated that knockout of some glycosylphosphatidylinositol (GPI) anchor biosynthesis components (e.g. *PIGA* and *PIGH*) restored glycoproteins to the cell surface under Tun stress. This suggests a novel interaction between N-glycosylation and GPI biosynthesis. RNAi in the *Drosophila* ER stress model confirmed that loss of the mannosyltransferase *Dpm1*, involved in mannose addition in several downstream glycosylation pathways, could rescue ER stress phenotypes. Testing these pathways (O- and C-mannosylation [*POMT2* and *DPY19L1*]), N-glycosylation [*ALG3*], GPI biosynthesis [*PIGM*]), we found loss of O-mannosylation and GPI biosynthesis partially recapitulates the *Dpm1* RNAi rescue of ER stress. These findings suggest GPI biosynthesis and/or mannose levels may underlie potential DPAGT1-CDG treatment options.

PrgmNr 3771 - A novel non-canonical splice site variant in *Delta 4-Desaturase, Sphingolipid 1 (DEGS1)* leads to loss of normal *DEGS1* splicing and results in functional deficiency of ceramide desaturase activity in two patients with hypomyelinating leukodystrophy

View session detail

Author Block: **C. Badduke**[1], J. Van Ziffle[1], A. M. Slavotinek[2], J. Tenney[2], A. Fay[3], H. Beale[4,5], A. Cheney[4], U. Hodoglugil[1], J. Y. Lee[6], P-M. Martin[7], N. J. Risch[7,8], J. D. Saba[6], T. Yip[2], O. M. Vaske[4,5], W. P. Devine[1]; [1]Dept. of Pathology, UCSF, San Francisco, CA, [2]Div. of Genetics, Dept. of Pediatrics, UCSF, San Francisco, CA, [3]Div. of Child Neurology, Dept. of Neurology, UCSF, San Francisco, CA, [4]Molecular, Cell and Dev.al Biology, UCSC, Santa Cruz, CA, [5]Genomics Inst., UCSC, Santa Cruz, CA, [6]Div. of Hematology/Oncology, Dept. of Pediatrics, UCSF, San Francisco, CA, [7]Inst. for Human Genetics, UCSF, San Francisco, CA, [8]Dept. of Epidemiology and Biostatistics, UCSF, San Francisco, CA
Disclosure Block: **C. Badduke:** None.
Delta 4-Desaturase, Sphingolipid 1 (DEGS1), at chromosome 1q42.11, encodes an enzyme that catalyzes the last step of the de novo ceramide biosynthetic pathway in which dihydroceramide (DhCer) is converted to ceramide (Cer). Variants in *DEGS1* have recently been associated with hypomyelinating leukodystrophy (MIM: 618404), a rare autosomal recessive condition that results in a complex neurological regressive disorder. To date, 14 homozygous or compound heterozygous variants have been reported (PMID: 30620337, 30620338, 31186544); 7/14 are truncating (4 nonsense, 3 frameshift), 7/14 are missense, and 2/14 (p.Asn255Ser and p.Trp107*) are recurrent. In this study, we report a novel homozygous intronic *DEGS1* indel at chromosomal position chr1:224378025-224378026 (hg19) detected by whole exome sequencing in 2 unrelated probands whose phenotypes overlap those reported in individuals with hypomyelinating leukodystrophy. Our variant, a non-canonical *DEGS1* splice site variant (c.825+4_825+5delAGinsTT; NM_003676.3), was classified according to ACMG guidelines as a variant of uncertain significance (VUS) as it is absent from population databases (gnomAD and dbSNP), has not previously been associated with disease (ClinVar and literature), and is not unanimously predicted by in silico algorithms to have a deleterious effect on splicing (Alamut). To characterize our VUS, we obtained a peripheral blood sample from one of our probands and performed transcriptome-wide RNA sequencing to study *DEGS1* splicing. *DEGS1* is expressed in whole blood at a similar level to that observed in multiple brain tissues (GTEx). Our results show that the c.825+4_825+5delAGinsTT variant affects the 5â splice donor sequence of exon 2 of *DEGS1* resulting in exon 2 skipping in the majority of transcripts detected in our proband. Other transcripts detected represent the use of various cryptic splice sites both upstream and downstream of the canonical exon 2 splice donor sequence. The resulting isoforms are predicted to be non-functional. Saturated and unsaturated sphingolipids were quantified by mass spectrometry in the plasma of the second proband and five pediatric controls. The ratios of C14, C16, C20, C22, C24, C24.1 and C26 dihydroceramide/ceramide were elevated (from 3-fold to 2,370-fold) in our proband compared to controls confirming functional deficiency of ceramide desaturase activity. In summary, we characterize a novel non-canonical *DEGS1* splice site variant using RNA sequencing and mass spectrometry and provide evidence to support its pathogenicity in hypomyelinating leukodystrophy. To our knowledge this is the first splice effect variant reported in *DEGS1*.

PrgmNr 3772 - An *in vivo* repurposing screen identifies novel therapeutic candidates for NGLY1 deficiency

View session detail

Author Block: **K. Hope**[1], R. Peterson[1], C. Y. Chow[2]; [1]Univ. of Utah Sch. of Med., Salt Lake City, UT, [2]Univ of Utah, Salt Lake City, UT
Disclosure Block: **K. Hope:** None.

Rare diseases impact more than 30 million Americans, however treatment options are limited or non-existent in approximately 90% of cases. NGLY1 deficiency, a rare disease with no effective treatment, is caused by autosomal recessive, loss-of-function mutations in the gene *N-glycanase 1* (*NGLY1*) and is characterized by global developmental delay, hypotonia, alacrima, and seizures. NGLY1 is a deglycosylase that removes GlcNAc sugar chains from glycoproteins that have been retrotranslocated from the endoplasmic reticulum (ER) lumen to the cytoplasm. Recent work demonstrated that the transcription factor NRF1 is translated into the ER membrane, glycosylated, retrotranslocated to the cytoplasm, and immediately degraded. Under proteasome stress, NRF1 is deglycosylated by NGLY1, allowing it to be cleaved and subsequently act as a transcription factor to upregulate proteasome genes. Unfortunately, therapeutic approaches targeting the known interaction between NGLY1 and NRF1 have yielded limited results. Here, we took an unbiased approach and conducted an *in vivo* small molecule screen to identify therapeutic compounds for NGLY1 deficiency. We used a *Drosophila melanogaster* model of NGLY1 deficiency that harbors a null allele of *dNGLY1* (*Drosophila NGLY1* ortholog) and displays lethality at the late pupal stage. In efforts to limit the time and cost associated with bringing a new drug to the NGLY1 deficiency population, we employed the Prestwick Chemical Library, consisting of 1280 off patent small molecules, 99% of which are approved by FDA/EMA regulatory agencies, to screen for compounds that rescue lethality in $dNGLY1^{KO}$ flies. We identified multiple compounds that modulate the serotonergic and dopaminergic signaling pathways, as well as iodine containing molecules, that rescued lethality in $dNGLY1^{KO}$ flies. Although serotonin and dopamine signaling impacts numerous cellular pathways, we hypothesize that serotonin and dopamine modulators rescue lethality in $dNGLY1^{KO}$ flies by inhibiting glycogen synthase kinase 3 (GSK3). Additionally, we utilized the Connectivity Map and identified a GSK3 inhibitor predicted to reverse gene expression changes in *dNGLY1* deficient flies, further supporting this hypothesis. GSK3 negatively regulates NRF1 activity, and disinhibiting NRF1 through GSK3 inhibition may rescue defects in $dNGLY1^{KO}$ flies. Additionally, iodine upregulates the transcription factor NRF2, which may compensate for the dysregulation of NRF1 upon the loss of NGLY1. This study demonstrates the power of *Drosophila* in therapeutic development for rare diseases, and similar small molecule screening strategies can be applied to other disorders.

PrgmNr 3773 - Epilepsy gene variants increase risk of seizure onset in familial Cerebral Cavernous Malformation

View session detail

Author Block: **S. Weinsheimer**[1,2], C. Fox[3], J. Nelson[1], A. Numis[3], M. Mabray[4], J. Zabramski[5], A. L. Akers[6], B. Hart[4], L. Morrison[7], C. E. McCulloch[8], H. Kim[1,2,8], Brain Vascular Malformation Consortium CCM Investigator Group; [1]UCSF Dept. of Anesthesia and Perioperative Care, Ctr. for Cerebrovascular Res., San Francisco, CA, [2]UCSF Inst. for Human Genetics, San Francisco, CA, [3]UCSF Dept. of Neurology, San Francisco, CA, [4]Univ. of New Mexico, Dept. of Radiology, Albuquerque, NM, [5]Barrow Neurological Inst., Dept. of Neurosurgery, Phoenix, AZ, [6]Angioma Alliance, Durham, NC, [7]Univ. of New Mexico, Dept. of Neurology, Albuquerque, NM, [8]UCSF Dept. of Epidemiology and Biostatistics, San Francisco, CA
Disclosure Block: **S. Weinsheimer:** None.
Background: Familial cerebral cavernous malformation (CCM) is an autosomal dominant disorder caused by mutations in *KRIT1*, *CCM2* and *PDCD10* genes. Seizures are a common presenting feature of CCM and clinical heterogeneity of familial CCM suggests a role for genetic modifiers. However, it is unknown whether epilepsy-related genes contribute to risk of seizure onset in familial CCM. We hypothesized that common genetic variants in known epilepsy or epilepsy-related genes are associated with earlier age of seizure onset in familial CCM patients. **Methods:** Familial CCM cases enrolled in the Brain Vascular Malformation Consortium were included (n=335). Time from birth to first seizure censored at last follow-up was assessed by medical records. Individuals were genotyped using the Affymetrix Axiom Genome-Wide LAT1 Human Array. We tested 28,670 single nucleotide polymorphisms (SNPs) mapping within 5kb of 935 genes identified as epilepsy-associated genes or genes related to epilepsy (reviewed in Wang et al, *Seizure* 2017) with minor allele frequency â¥1%. With interval censored survival analysis assuming a Weibull distribution, we tested whether the number of minor alleles was associated with age of first seizure onset, adjusting for age at enrollment and sex. We report hazard ratio (HR) and 95% confidence interval (CI), and significance was based on Bonferroni adjustment for multiple comparisons (P=0.05/28,670 SNPs =1.74E-06).**Results:** Overall, 139 (41.5%) of cases had seizure over a mean analysis time of 33.5Â±22.3 years (11,860 total years). Genetic variants mapping to introns in three genes were significantly associated with seizure onset when adjusting for age and sex. NEDD4 like E3 ubiquitin protein ligase (*NEDD4L*, rs17064791 G>A) stratified seizure risk, with carriers of the minor allele at nearly 5-fold increased risk of earlier seizure onset compared to non-carriers (HR=4.8; CI: 2.6-8.6; P=1.7E-07). In addition, genetic variants in leucine rich repeat kinase 2 (*LRRK2*, rs73106354 A>G, HR=4.4; CI: 2.5-7.7; P=3.3E-07) and erb-b2 receptor tyrosine kinase 4 (*ERBB4*, rs72945032 T>C, HR=2.6; CI: 1.8-3.9; P=9.4E-07) also significantly increased risk of earlier seizure onset.**Conclusion:** Genetic variants in the epilepsy-associated gene *NEDD4L*, and epilepsy-related genes *LRRK2* and *ERBB4* were associated with earlier seizure onset in familial CCM patients, lending further evidence that these genes may be relevant for familial CCM disease severity. Further work is needed to confirm these findings in a replication cohort. The identification of individuals carrying epilepsy genetic variants may aid in CCM patient stratification for more personalized treatment.

PrgmNr 3774 - Genetic modifiers of NGLY1 deficiency identified through a *Drosophila* genetic screen point to the role of NGLY1 in ERAD

View session detail

Author Block: **T. Tu'ifua**[1], D. Talsness[2], K. Owings[2], C. Y. Chow[3]; [1]Dept. of Human Genetics, Univ. of Utah, Salt Lake City, UT, [2]Univ. of Utah, Salt Lake City, UT, [3]Univ of Utah, Salt Lake City, UT
Disclosure Block: **T. Tu'ifua:** None.

N-glycanase 1 (NGLY1) deficiency is a rare disease caused by autosomal recessive loss of function mutations in the *NGLY1* gene and is the only known congenital disorder of deglycosylation. Patients suffer from movement disorder, developmental delay, liver dysfunction, and alacrima. NGLY1 removes N-linked glycans from glycoproteins in the cytoplasm and is thought to help clear misfolded glycoproteins from the endoplasmic reticulum (ER) through the ER associated degradation (ERAD) pathway. Despite this, NGLY1â□□s physiological significance in ERAD is not understood. One way to understand disease pathogenesis is to investigate the effects of genetic variation and modifier genes on disease presentation. Our lab created a *Drosophila* model of NGLY1 deficiency that faithfully recapitulates several disease phenotypes observed in patients, including movement disorder, seizures, and lethality. We crossed this *Drosophila* model of NGLY1 deficiency with the *Drosophila* Genetics Reference Panel (DGRP), a collection of ~200 inbred fly strains with fully sequenced genomes, and scored for proportion of *NGLY1* knockdown flies surviving to adulthood. The genetic variation of the DGRP led to a spectrum of lethality ranging from strains that had 100% lethality to 100% viability. A genome wide association study (GWAS) generated a list of 61 candidate modifier genes of NGLY1 deficiency. Nine of these candidate genes encode ER resident proteins, proteins with known ERAD functions, or are involved with protein homeostasis. *CG31690* and *CG4341* were two independent unlinked hits in the GWAS that are orthologs of a single human gene, *TMTC2*, which encodes an ER adaptor protein involved with calcium homeostasis. *Hrd3* and *CG8405*, orthologs of *SEL1L* and *TMEM259*, were both hits and are known components of ERAD complexes that retrotranslocate misfolded proteins from the ER to the cytoplasm for degradation. We are functionally characterizing several of these candidate modifier genes of NGLY1 deficiency using *Drosophila* and cell culture models. The study of modifier genes can provide new insights into the etiology of the disease and functions of NGLY1, laying the foundation for the development of personalized therapeutics.

PrgmNr 3775 - Human genetics is great: Bone Mineral Density Studies in Patients with IBMPFD using DEXA

View session detail

Author Block: **R. Columbres**[1,2], M. Nguyen[1], V. Kimonis[1]; [1]Univ. of California, Irvine, Irvine, CA, [2]William Carey Univ. Coll. of Osteopathic Med., Hattiesburg, MS
Disclosure Block: **R. Columbres:** None.
Background: Inclusion Body Myopathy (IBM), associated with combinations of Pagetâ□□s Disease of the Bone (PDB), Frontotemporal Dementia (FTD) (IBMPFD), and amyotrophic lateral sclerosis (ALS) is a rare, autosomal dominant, adult-onset, progressive, and multisystem disorder caused by mutations in *VCP* gene. The disorder is characterized by progressive axial and proximal muscle weakness present in 90% starting in their thirties, Paget disease in 42% starting in their thirties, and FTD manifestations such as behavioral, cognitive and language impairment starting in their fifties, which is present in approximately 30% of individuals. Dual-energy X-ray absorptiometry (DEXA) is a gold standard for body composition measurements. It generates values for bone mineral densities (BMD), lean mass, and fat mass. The goal of this cross-sectional study of patients suffering from IBMPFD is to accurately detect differences in bone mineral density (BMD), bone mass %, lean mass %, and fat mass %: 1) at different stages of the myopathic disease 2) in subjects who additionally had Paget disease of bone, 3) to establish correlations between lean mass and bone mass.
Results and Conclusion: Our cohort included 19 subjects with IBMPFD (11 males, 8 females), three carriers, and six controls. Twelve (63.2%) had myopathy without Paget at a mean age of 53.8 Â± 9.4 (mean Â± SD), and mean disease onset of 43.3 Â± 9.1 (mean Â± SD). Seven (36.8%) had myopathy with Paget at a mean age of 49.0 Â± 9.2 (mean Â± SD), and mean disease onset of 39.2 Â± 6.9 (mean Â± SD). Within the myopathy-only group, 4 had osteopenia, and one had osteoporosis. One individual from each carrier, control, and myopathy with Paget groups also had osteopenia. We observed a significant inverse relationship between age and duration of disease with left hip T-score in the myopathy-only group. We found a statistically significant difference in the left hip scan BMD between myopathy-only and control groups. The myopathy-only group exhibited reduced lean mass % and increased total fat % compared to the control group. The myopathy with Paget group had increased T-scores and Z-scores in lumbar (L1-4) vertebrae compared to findings in other groups. Overall, DEXA is a useful method of evaluating BMD, total fat mass, and lean mass estimation in individuals with IBMPFD. Within the myopathy-only group, 33.3% had osteopenia, and 8.3% had osteoporosis. Interestingly, DEXA in individuals with Paget disease of bone showed increased density in the lumbar spine, a common area for Paget. DEXA showed that there was a reduction in lean mass and increasing fat with increasing age. A larger cohort would be needed to show its value in monitoring the progression of the disease.

PrgmNr 3776 - Identification of small and large variants in familial amyotrophic lateral sclerosis

View session detail

Author Block: **S. Chan Moi Fat**[1], E. McCann[1], L. Henden[1], K. Williams[1], D. Rowe[1,2], J. Fifita[1], I. Blair[1]; [1]Macquarie Univ. Ctr. for Motor Neuron Disease Res., Dept. of BioMed. Sci., Faculty of Med. and Hlth.Sci., Sydney, Australia, [2]Dept. of Clinical Med., Faculty of Med. and Hlth.Sci., Macquarie Univ., Sydney, Australia
Disclosure Block: **S. Chan Moi Fat:** None.
Amyotrophic lateral sclerosis (ALS; also known as Lou Gehrigâ□□s disease) is a devastating neurodegenerative disease, caused by the progressive loss of motor neurons in the brain and the spinal cord. To date, gene mutations remain the only proven cause of ALS, with more than 30 ALS genes identified. However, 90% of ALS cases (about 40% of familial and 90% of sporadic ALS) still have an unknown genetic cause. Most research in ALS has focused on small nucleotide level variants only. This project extends ALS gene discovery to also consider large structural variation by using bioinformatic pipelines to analyse whole genome sequencing (WGS) data for the identification of both small and large genetic alterations in a small ALS family that has limited power for traditional gene discovery methods.

WGS was performed on two affected individuals from a family with a history of ALS (MQ52). To identify single nucleotide variants and small insertion/deletions potentially acting as the pathogenic gene mutation causing ALS within this family, shared variant analysis using custom bioinformatics was carried out to filter for novel nonsynonymous and indel variants absent or extremely rare from publicly available control databases. Variants were validated using the Integrative Genomics Viewer and direct sequencing. To identify structural variants (SVs), we have developed a comprehensive pipeline which calls SVs from WGS data using multiple tools such as Manta, Lumpy and MetaSV. The tool Duphold, is then used to filter SVs based on quality scores, followed by SV annotation using Reciprocal Overlap Annotator and AnnotSV to generate a list of shared or overlapping between the two affected individuals in the family.

Initial analysis of family MQ52 identified a total of 14 small candidate variants (13 single nucleotide variants and one indel). Using the SV discovery pipeline, 80 novel/rare SVs (from a total of over 8000 total family SVs) were identified as potential causes of disease in this family, out of which 44 SVs lie in coding regions of the genome. This SV pipeline is still undergoing refinement and following further filtering resultant SVs will be validated in the lab using established PCR techniques.

Our gene discovery pipelines can successfully identify both small and large variants with the potential to cause disease within small ALS families. Identifying novel ALS variants has great potential to identify novel ALS genes and pathways, that could help with the development of future treatments. Given that SVs have been largely understudied in ALS, finding ALS-linked SVs could help to elucidate some of the missing genetic causes of ALS.

PrgmNr 3777 - Identification of X-linked missense variants in *TAF1* in 4 unrelated families with autism spectrum disorder (ASD)

View session detail

Author Block: **J. Lim**[1], Y. Qiao[2], K. Calli[2], S. Martell[2], S. J. M. Jones[3], S. W. Scherer[4,5], S. M. Lewis[2]; [1]Faculty of Med., Univ. of British Columbia, Vancouver, BC, Canada, [2]Med. Genetics, Univ. of British Columbia, Vancouver, BC, Canada, [3]Michael Smith Genome Sci. Ctr., Vancouver, BC, Canada, [4]The Hosp. for Sick Children, Toronto, ON, Canada, [5]MSSNG (research.mss.ng), Toronto, ON, Canada
Disclosure Block: **J. Lim:** None.
TAF1 (TATA-Box Binding Protein Associated Factor 1) is an X-linked gene that encodes the largest component and core scaffold of the TFIID basal transcription factor complex. It plays an important role in neurodegenerative diseases and developmental delay. Recently, an increasing number of cases have been reported with the variants in this gene grouped as a new neurodevelopmental syndrome (TAF1/MRXS33 intellectual disability syndrome). The common clinical features appear early in life with hypotonia, facial dysmorphia, developmental delay, intellectual disability (ID) and/or autism spectrum disorder (ASD). We identified four X-linked variants in this gene (p.Asp1496Ala, p.Pro215Leu, p.Leu619Phe, and p.Gly1391Arg) in four boys from four unrelated Asian families (one of these is a mixed heritage Asian-white family). All of the boys are affected with autism spectrum disorder (ASD). Out of the four probands, one of them showed a more severe phenotype (Complex-5) while the other three had either Simple-1 or Simple-2 according to our phenotype semi-quantitative evaluation criteria. Interestingly, we noticed that the genomic location of the variant in the most severely affected patient was located in the hot region of the gene. This area is enriched with a higher number of reported pathogenic/likely pathogenic variants. A detailed genotype-phenotype correlation analysis will be summarized among our patients and published cases in the future. Finally, other variants in multiple strong ASD candidate genes were found in each of our cases, inherited from both parents. Our data suggest that different types and locations of the variants in *TAF1* genes, as well as the additive or epistatic roles of other variants from different ASD candidate genes might contribute to the diversity of phenotypes in this syndrome.

PrgmNr 3778 - Loss of function variants in DIP2C are associated with neurocognitive delays

View session detail

Author Block: **S. Htun**[1], T. Ha[1], L. Pasquier[2], C. Zweier[3], A. Reis[4], V. Siu[5], B. DeVries[6], R. Pfundt[6], M. Nizon[7], M. Vincent[7], B. CognÃ©[7], U. Nawaz[8], I. Voineagu[9], A. Slavotinek[1]; [1]Div. of Med. Genetics, Dept. of Pediatrics, Univ. of California San Francisco, San Francisco, CA, [2]Service de GÃ©nÃ©tique MÃ©dicale, HÃ´pital Sud, Rennes, France, [3]Dept. of Human Genetics, Inselspital, Bern Univ. Hosp., Univ. of Bern, Bern, Switzerland, [4]Inst. of Human Genetics, Friedrich-Alexander-Univ. Erlangen-NÃ¼rnberg, Erlangen, Germany, [5]London Hlth.Sci. Ctr. and Dept. of Pediatrics, Schulich Sch. of Med. and Dentistry Western Univ., London, ON, Canada, [6]Dept. of Human Genetics, Radboud Univ. Med. Ctr. and Donders Inst. for Brain, Cognition and Behavior, Nijmegen, Netherlands, [7]CHU Nantes, Service de GÃ©nÃ©tique MÃ©dicale, L'institut du thorax, Univ. Nantes, Nantes, France, [8]Adelaide Med. Sch. and Robinson Res. Inst., The Univ. of Adelaide, Adelaide, Australia, [9]Sch. of Biotechnology and Biomolecular Sci., Univ. of New South Wales, Sydney, Australia

Disclosure Block: **S. Htun:** None.

The Disconnected (disco)-interacting protein 2 (DIP2) homologues, *DIP2A*, *DIP2B*, and *DIP2C*, are highly conserved and broadly expressed in the central nervous system of vertebrates. We present 5 females and 3 males with heterozygous variants predicting loss of function (LOF) in *DIP2C*. Four patients had *de novo* variants, one child had a nonsense variant inherited from a father with speech delays and two siblings had a nonsense variant inherited from a mother with speech delays. Patients were evaluated at 5-18 years and all manifest developmental delays, with ages of walking ranging from normal to 2.5 years and ages at first word from 13 months to 2 years. Four out of 8 patients had speech limited to single words at 5-10 years of age and attention deficit hyperactivity disorder (ADHD; 3/8), sensory integration disorder (1/8), hypotonia (2/8) and autism (1/8) were noted. Additional anomalies were relatively rare but included a triangular facial shape with a prominent forehead, thin hair with a high anterior hairline, strabismus and hearing loss. Deletion of *DIP2C* has repeatedly been observed with deletion of *ZMYND11* in 10p15.3 microdeletion syndrome that is characterized by global developmental delays, behavioral disturbances, dysmorphic features, brain anomalies, seizures and short stature. Although *ZMYND11* was considered the critical gene for this phenotype, mosaicism for a 67.6 kb deletion impacting *DIP2C* was reported in a patient with cerebral palsy and ADHD who was reported to be a â□□slow learnerâ□□, and a *de novo* frameshift variant in *DIP2C* was identified in a patient with autism. *Dip2c* was expressed in the brain of postnatal mice, and Dip2c$^{Î□2bp}$ mice that were homozygous for a two base pair deletion showed altered expression of genes associated with memory and neuropeptide signaling. We used the BrainSpan Atlas of the developing human brain to examine the expression of *DIP2C* from 8 post-conception weeks (pcw) to 40 years in neocortex and non-neocortex. Expression was highest in neocortex and non-neocortex from 10-24 pcw, although expression was present at all developmental stages. Cell-type specific enrichment analysis of *DIP2C* showed expression in all brain tissue subtypes. We targeted the zebrafish orthologues *dip2ca* and *dip2cb* using CRISPR/Cas9 but did not observe larvae with consistent morphological differences compared to controls although experiments to assess behavior in larvae with stable knockdown are ongoing. The loss of function variants in these patients and brain expression pattern of *DIP2C* support a role for this gene in neurocognitive function, although these families demonstrate that LOF variants may be inherited from parents with milder phenotypes.

PrgmNr 3779 - A mouse model recapitulating Bruck Syndrome provides insight into the role of *Plod2* in bone and cartilage

View session detail

Author Block: **A. Kot**[1], I. Duran[1,2], J. H. Martin[1], D. Wachtell[1], M. Weis[3], D. R. Eyre[3], J. Zieba[1], D. Krakow[1]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Univ. of Malaga, Malaga, Spain, [3]Univ. of Washington, Seattle, WA

Disclosure Block: **A. Kot:** None.

Bruck Syndrome is an autosomal recessive form of osteogenesis imperfecta characterized by bone fragility, short stature, and joint contractures resultant from biallelic mutations in either the genes encoding *PLOD2* or *FKBP10*. *PLOD2* encodes LH2 which is a component of an endoplasmic reticulum (ER) complex involved in type-I collagen telopeptide hydroxylation. Bruck Syndrome has features distinct from other forms of osteogenesis imperfecta, such as joint contractures, with pathophysiology that is not well understood. Studies on a previously published *Plod2* knockout mouse were limited due to early embryonic lethality. To study the pathophysiology of musculoskeletal development in Bruck Syndrome, we generated a mouse model with a homozygous mutation identified in two independent families with recurrent fractures, scoliosis, and congenital contractures. The mutation, *Plod2* c.1559dupC predicting the amino acid change p.Val523Cysfs*7, was introduced using targeted homologous recombination. *Plod2* is alternatively spliced and translated into short (LH2a) or long (LH2b) isoforms. The c.1559dupC mutation is in exon 13A, specific to LH2b, and is predicted to lead to loss of the long form of the protein. *Plod2*$^{V523fs/V523fs}$ mice did not survive past P0. At E18.5, *Plod2*$^{V523fs/V523fs}$ mice were smaller, had bilateral forelimb and hindlimb contractures, and absent cervical spine curvature. LH2b mRNA was near absent in *Plod2*$^{V523fs/V523fs}$ mouse calvaria. LH2 protein levels in *Plod2*$^{V523fs/V523fs}$ osteoblasts were also decreased compared to wild type. Type I collagen C-telopeptide lysine residue hydroxylation was significantly reduced in *Plod2*$^{V523fs/V523fs}$ bone suggesting that crosslinking is negatively impacted, contributing to impaired bone properties. Compared to wild type, *Plod2*$^{V523fs/V523fs}$ growth plates showed increased length of the hypertrophic zone, increased hypertrophic chondrocyte volume, and diminished collagen production as detected by picrosirius staining suggesting a role for *Plod2* in chondrocyte development. This mouse model will aid in determining the function of *Plod2* in the musculoskeletal system and help uncover the molecular mechanisms underlying Bruck Syndrome.

PrgmNr 3780 - Expanding clinical and molecular evidence of *CYP26B1* involvement in retinoic acid catabolism defects

View session detail

Author Block: **K. C. Silveira**[1], I. Chacon Fonseca[2], O. ArtigalÃis[3], A. Iacovone[4], E. Campos[5,6], D. P. Cavalcanti[4], P. Kannu[1]; [1]Dept. of Med. Genetics, Univ. of Alberta, Edmonton, AB, Canada, [2]Clinical Genetics, Dept. of Oncology, Lakeridge Hosp., Oshawa, ON, Canada, [3]Clinical Genetics Unit, Children's Hosp., Grupo Hosp.ar ConceiÃ§Ã£o, Porto Alegre, Brazil, [4]Skeletal Dysplasia Group, Univ. of Campinas, Campinas, Brazil, [5]Genetics & Genome Biology program, Hosp. for Sick Children, Toronto, ON, Canada, [6]Dept. of Molecular Genetics, Univ. of Toronto, Toronto, ON, Canada

Disclosure Block: **K.C. Silveira:** None.

Homozygous missense variants in *CYP26B1* have been described in very few individuals with skeletal abnormalities. The P450 cytochrome CYP26B1 metabolizes retinoic acid (RA) in the developing embryo to tightly control RA levels and locally regulates signaling. The *CYP26B1* mutation phenotype ranges from a severe lethal presentation with skull defects, craniosynostosis, encephalocele, radiohumeral fusion, oligodactyly and a narrow thorax to a milder presentation described in an adult affected by craniosynostosis, radiohumeral joint limitation, hearing loss and intellectual disability. Here, we describe three novel *CYP26B1* variants in two unrelated families. Family 1 with two siblings affected by arachnodactyly, reduced radio-ulnar joint movement, conductive hearing loss and a mild to moderate learning disability. Craniosynostosis was not present in either sibling. In a second unrelated and consanguineous family, a lethal presentation in a female stillbirth with large hydrocephalus associated with generalized spina bifida occulta and poor mineralization of the whole skeleton and limb defects including oligodactyly was found. Whole exome sequencing of family 1 revealed two novel CYP26B1 variants in trans: c.353C>T (p.Pro118Leu) and c.701G>A (p. Arg234Gln). We hypothesize that these variants generate a protein with reduced enzyme activity and the degree of retained enzymatic activity drives phenotypic severity. However, further studies are still necessary to prove their pathogenicity. These siblings represent a milder *CYP26B1* phenotype when compared to previously described cases. In family 2, a novel homozygous point variant (c.1083C>A) was identified through Sanger sequencing of *CYP26B1*. In silico analysis suggested that the variant activates a cryptic splice site altering splicing. A minigene assay showed the variant leads to a deletion of part of exon 5 generating mRNA with a deletion of 21 amino acids (p.Val361_Asp382del). This deletion includes the ExxR motif and SRS-5 region, which are related to the stabilization of the meander loop and the maintenance of the CYP tertiary structure and the orientation of the substrate near the heme center, respectively. This deletion leads to the most severe clinical presentation so far. Taken together, we propose the phenotype associated with *CYP26B1* varies depending on the type of variant, although all are inherited in a recessive matter. Furthermore, we describe here the first patient associated with a splicing variant resulting in a severe and lethal phenotype. With their description, we add to the genotypic and phenotypic spectrum seen with defects of the catabolism of RA.

PrgmNr 3781 - FDXR Mitochondriopathy: The challenges of reclassifying variants of unknown significance in a newly described syndrome

View session detail

Author Block: **A. Lee**[1], L. Forero[2], J. Lenberg[3], M. Wright[4], L. Guidugli[4], R. R. Mardach[5]; [1]Univ. of California, San Diego, San Diego, CA, [23]Rady Children's Hosp., San Diego, CA, [4]Rady Children's Inst. for Genomic Med., San Diego, CA, [5]Rady Children's Hosp. of San Diego, San Diego, CA
Disclosure Block: **A. Lee:** None.
Recently emerging genetic technologies such as whole genome sequencing (WGS) have begun to uncover new and uncharted genetic changes, many of which have associations with previously described phenotypes. We present two affected family members found to have biallelic compound heterozygous missense Variants of Uncertain Significance (VUS) in FDXR identified by genome-wide sequencing. FDXR is associated with autosomal recessive Auditory neuropathy and optic atrophy (MIM: # 617717), but emerging evidence suggests an association with a yet unnamed mitochondriopathy syndrome encompassing a wide range of severity, which we have designated FDXR Related Neurodegenerative Infantile Optic Atrophy. Evidence for gene-disease association includes a cohort of over 20 individuals harboring missense variants, murine studies, in-vitro studies, and supports genotype-phenotype correlations. Notably, previously reported missense variants are located in regions of the gene where our probands' missense variants are also located, suggesting a potential important functional region. Variant reclassification is important for the genetic and medical community at large but also to patients who may not be able to access necessary surveillance, therapy, insurance coverage for further testing, supportive services, or parental prenatal or preimplantation testing without a definitive diagnosis. The ACMG has set forth evidence-based criteria for variant classification, but generating the evidence needed for reclassification in a timely manner is a difficult feat. Here, we provide additional evidence and an official designation for a newly described and rare syndrome involving variants in the FDXR gene. In addition, we detail steps taken towards providing a diagnosis for this family and expand upon the overall process of reclassification as an essential and impactful scenario which will continue to be increasingly common as access to genetic sequencing increases.

PrgmNr 3782 - Long-Chain Fatty Acid Oxidation Disorder Gene Variants Identified Through a Gene Panel Sponsored Program from Clinically Confirmed and Suspected LC-FAOD

View session detail

Author Block: **V. Rangel Miller**[1,2], T. Chettiath[2], B. A. Johnson[3], D. L. Marsden[2], J. L. Merritt[2], S. Sarafrazi[2], A. Willcock[3], N. Miller[4]; [1]Ultragenyx, Kirkland, WA, [2]Ultragenyx Pharmaceutical Inc., Novato, CA, [3]Invitae Corp., San Francisco, CA, [4]Ultragenyx, Brisbane, CA
Disclosure Block: **V. Rangel Miller:** Salary/Employment; Ultragenyx Pharmaceutical Inc.
Long-chain fatty acid oxidation disorders (LC-FAOD) are rare autosomal recessive conditions caused by defects in genes encoding mitochondrial enzymes that convert long-chain fatty acids into energy. Patients with LC-FAOD may have characteristic elevations in plasma acylcarnitine profiles, identified through newborn screening follow-up and through testing patients with ongoing clinical signs and symptoms. **Methods**: Patients in the US, Canada, and Mexico of all ages who have either a clinical diagnosis or suspicion of LC-FAOD with a confirmatory acylcarnitine test either ordered or performed are eligible for this no-charge genetic testing program. The next generation sequencing gene panel with copy number variant (CNV) detection includes the 6 genes associated with LC-FAOD (*ACADVL, CPT1A, CPT2, HADHA, HADHB, SLC25A20*) plus 18 additional genes associated with disorders that cause a similar acylcarnitine profile. **Results**: 283 patients have been tested as of May 12, 2021. Newborn screening results provided for 198 patients were as follows: 144 (73%) positive, 32 (16%) unknown, 22 (11%) negative; and confirmatory results provided for 165 patients were as follows: 39 (24%) positive, 98 (59%) inconclusive, 28 (17%) negative. Clinical signs reported for 82 patients demonstrated rhabdomyolysis and myopathy as most common among adolescents and adults (n= 46), while hypoketotic hypoglycemia and elevated creatine kinase were most common among younger patients (n= 36). Test results identified one or more LC-FAOD gene variant in 110 patients, including 54 variants of uncertain significance (VUS), 7 likely pathogenic (LP), and 41 pathogenic (P) variants. Results provided a positive diagnosis (P/P, LP/P, LP/LP) for 14 patients, and potential positive diagnosis (LP/VUS, P/VUS) for 16 patients. Variants were found in 14 additional non-LC-FAOD genes on the panel (*ACADM, ACADS, ACADSB, ETFA, ETFB, ETFDH, HADH, HMGCL, HMGCS2, MLYCD, NADK2, SLC22A5, SLC52A1, SLC52A3*), including 36 VUS, 6 LP, and 30 P. 4 patients had heterozygous variants in two or more LC-FAOD genes and 17 patients had heterozygous variants in a LC-FAOD gene and one or more non-LC-FAOD genes. 61 patients had only 1 LC-FAOD gene variant identified and 119 patients had no variants identified in any gene on the panel. **Conclusion**: Program results demonstrate the diverse composition of gene variants in this patient cohort and suggest the need for further approaches to resolve VUS variants and identify previously undetected variants in patients with suspected LC-FAOD.

PrgmNr 3783 - Novel heterozygous *OPA3* mutation in a family with congenital cataracts, sensorineural hearing loss and neuropathy, without optic atrophy

View session detail

Author Block: **M. Penon Portmann**[1], A. J. Kruse[2], J. So[1]; [1]Univ. of California, San Francisco, San Francisco, CA, [2]Univ. of Washington, Seattle, WA
Disclosure Block: **M. Penon Portmann:** None.
Autosomal dominant optic atrophy is caused by heterozygous pathogenic variants in *OPA3* (OMIM #165300), and is characterized by optic atrophy with cataracts, peripheral neuropathy and hearing loss (PMID: 1119193, 28050599). *OPA3* encodes a mitochondrial protein likely involved in the regulation of mitochondrial fission. To date, only ten pathogenic variants in *OPA3* have been reported. Here, we characterize the phenotype in a family with a novel variant in *OPA3* and notably without optic atrophy.Two probands with multiple affected relatives were assessed. Proband 1 is a 71-year-old male with bilateral congenital cataracts and bilateral sensorineural hearing loss, as well as tremor and distal sensorimotor polyneuropathy on electromyography (EMG) and nerve conduction studies (NCS). Proband 2 is the 27-year-old daughter of proband 1 with a history of bilateral sensorineural hearing loss, congenital bilateral cataracts, paresthesias in both distal lower extremities and tremor. Family history is remarkable for an autosomal dominant pattern of inheritance with multiple similarly affected family members.Clinical exome showed variants of uncertain significance (VUS) in *OPA3*, *DMPK*, *MYH14*, *TECTA* and *TTN*. Only the variants in *OPA3* (NM_025136:c.30G>C, p.Lys10Asn) and *DMPK* (NM_001081563:c.91G>C, p.Gly31Arg) segregated with 5 affected family members and were absent from 2 unaffected family members. The missense variant in *OPA3* is absent from gnomAD, with damaging *in silico* predictions, and occurring in a highly conserved residue, yet is not a previously reported pathogenic variant.The familyâ□□s clinical presentation has significant phenotypic overlap with previously reported cases of *OPA3*-related disorder, except for a notable lack of optic atrophy. Given the *DMPK* variant also tracks with affected individuals in the family, we cannot dismiss its significance, though no family members are reported to have myotonia or weakness, EMG/NCS in proband 1 was inconsistent with a myopathy and *DMPK* missense variants are not known to be associated with myotonic dystrophy. Our case broadens the clinical and genetic spectrum associated with *OPA3* mutations and highlights that optic atrophy is not an obligate feature of OPA3-related disorders.

PrgmNr 3785 - Recessive genetic risk factor for hemochromatosis is associated with damaging iron deposition in motor circuits of the human brain

View session detail

Author Block: **R. Loughnan**, J. Ahern, C. Tompkins, C. E. Palmer, W. Thompson, T. Jernigan, A. Dale, M. Boyle, C. Chieh Fan; Univ. of California, San Diego, La Jolla, CA
Disclosure Block: **R. Loughnan:** None.
Hereditary hemochromatosis (HH) is an autosomal recessive genetic disorder that leads to iron overload in the body causing oxidative damage to affected organs. HH type 1 appears to be predominantly associated with homozygosity for the mutation p.C282Y - with an odds ratio of 411 in men. Previous case studies have reported tentative links between HH and movement disorders, e.g. Parkinson's disease, and some have found abnormalities on brain magnetic resonance imaging (MRI) scans in regions related to movement (basal ganglia). Here we sought to investigate the impact of p.C282Y homozygosity by i) associating T2 intensity differences across all imaging voxels of the human brain (lower intensities indicate higher iron deposition) and ii) their associations with movement disorders within UK Biobank. For those with neuroimaging data (154 p.C282Y homozygotes, 595 matched controls), we found p.C282Y homozygosity was associated with signal profile changes that are consistent with substantial iron deposition in motor circuits of the human brain (basal ganglia, red nucleus and cerebellum; Cohen's d > 1 for differences in T2 intensities). Across the whole UK Biobank sample with qualified data (5,568 p.C282Y homozygotes, 496,968 controls), we found a significant enrichment for movement disorders in male homozygotes (OR (95% CI) = 1.82 (1.27-2.61), p=0.001), but not females (OR (95% CI) = 1.10 (0.69-1.78), p=0.71). This result is consistent with previous findings of higher disease burden for p.C282Y homozygote males vs females. Among the 31 p.C282Y homozygote males with a movement disorder only 7 had a concurrent HH diagnosis. This is of importance given the differences in treatments for HH vs movement disorders like Parkinsonâ□□s disorder. These findings suggest considering p.C282Y homozygosity as a risk factor for movement disorders in males.

PrgmNr 3786 - RNA-sequencing of tissue from family members identifies causal mutation in children

View session detail

Author Block: **R. Wang**[1], J. Woods[2], F. Barthelemy[3], E. D. DOUINE[4], C. Miceli[1], S. F. Nelson[5]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]Universit of California Los Angeles, Los Angeles, CA, [3]615 charles e young, los angeles, CA, [4]UCLA, Los Angeles, CA, [5]UCLA Med Ctr, Los Angeles, CA
Disclosure Block: **R. Wang:** None.
Duchenne muscular dystrophy (DMD, OMIM #310200) is a rare X-linked pediatric disorder characterized by progressive muscle weakness. The incidence is about 1:3,500 male births, making DMD one of the most common serious genetic diseases of childhood. DMD follows a well-defined progression beginning with delayed walking and muscle weakness in early childhood followed by loss of ambulation and full time wheelchair use around age 10. Death occurs by the second or third decade usually resulting from cardiorespiratory failure. The genetic cause of DMD is a defect in the massive 2.2Mb *DMD* gene which encodes the structural protein dystrophin. The most common mutations are exonic deletions (72%), duplications (10%), indels (5%), splice site mutations (3%) and point mutations (9%). Still, 2-5% of cases lack a molecular diagnosis. We present two cases of DMD - one with a known cause and the other unsolved despite standard genetic testing using multiplex ligation-dependent probe amplification and Sanger sequencing of *DMD* exons - where we identified the causal mutations by performing RNA-seq on needle biopsy samples taken from their mothers. In the case of the previously known mutation, we clearly observed a 1bp deletion in *DMD* exon 30 in roughly half the reads of the mother indicating biallelic expression across the muscle RNA-seq as expected. Since this is an X-linked gene subjected to X-inactivation, we are effectively observing the mixture of mutant and non-mutant DMD transcripts within the maternal muscle biopsy. This result matched the gene test results from a clinical lab on the affected son. The second case had strong suspicion of muscular dystrophy based on clinical course and high serum creative kinase. However, clinical testing was inconclusive, and by the time of evaluation, the patient had such severe muscle wasting in his 20s that no muscle biopsy could be reasonably obtained. RNA-seq analysis of the muscle biopsy from the mother identified a tandem duplication of the region between intron 44 and intron 63 of the *DMD* gene, which was confirmed as the causal DNA mutation (duplication of exons 45-63) in her son. Both carrier moms showed heterozygous expression of the mutation at the RNA level. In this context, using physiologically relevant tissue from family members enables a search by RNA-seq to identify mutations that affect mRNA and may be otherwise cryptic from DNA studies.

PrgmNr 3787 - Transcriptomic profiling of paired skeletal muscles from healthy donors using bulk RNAseq and snRNAseq to identify DMD susceptibility factors

View session detail

Author Block: **S. Nieves-Rodriguez**[1], F. Barthelemy[2], J. Woods[3], E. D. DOUINE[4], R. T. Wang[5], D. Scripture-Adams[1], K. Chesmore[1], M. Miceli[1], S. F. Nelson[6]; [1]Univ. of California, Los Angeles, Los Angeles, CA, [2]615 charles e young, los angeles, CA, [3]Valley Children's Hosp., Madera, CA, [4]UCLA, Los Angeles, CA, [5]Univ California los Angeles, Los Angeles, CA, [6]UCLA Med Ctr, Los Angeles, CA
Disclosure Block: **S. Nieves-Rodriguez:** None.
In Duchenne Muscular Dystrophy (DMD), the absence of dystrophin results in a progressive and irreversible loss of skeletal muscle causing loss of ambulation (LOA). The age at LOA can be variable between patients independent of the causal variant, suggesting the existence of environmental and/or genetic disease modifiers. Furthermore, the disease process is characterized by differential progression of muscle groups indicating specific, but not yet elucidated, mechanisms of myofiber susceptibility of different muscles. Various muscular dystrophies have differential muscle group susceptibility. In DMD, vastus lateralis (VL) is affected earlier in the disease process relative to the tibialis anterior (TA), which is substantially relatively protected for years. Thus, querying for genes with differences at the whole tissue and/or single cell level between these differentially susceptible muscles should provide insights into epigenetic mechanisms in muscle development, and provide insights into disease susceptibility mechanisms more broadly. To observe differences between muscle groups, we performed paired core needle biopsies of TA and VL in healthy young adults (n=11). Using bulk RNA sequencing (RNAseq), we found that 2299 (of 18,254) genes are differentially expressed between paired healthy VL and TA. Strikingly supportive of a functionally relevant role in protection/susceptibility to damage from loss of dystrophin, 10 of 14 (71%) genes that are described DMD modifiers were identified among these genes. We found relatively higher expression of extracellular matrix (ECM) genes and lower expression of ribosomal processes genes in TA. Using single nuclei RNAseq from frozen tissue, we found different cell type proportions between TA and VL (n=3). Genes higher in TA in bulk RNAseq are localized to expression from fibroblasts, pericytes, endothelial cells and smooth muscle, and those higher in VL are more expressed from myofibers nuclei. Our findings suggest that the cell types that predominate in the production of ECM, such as fibroblasts (more abundant in TA), may be an important component of intrinsic muscle protection to dystrophy. Understanding molecular signatures underlying phenotypical heterogeneity across muscles is of importance to: i) better understand the inter- and intra-individual heterogeneity observed in muscular dystrophies, ii) develop treatments to mimic the intrinsic protection against disease observed in typically less affected muscles, with the goal of using them as complements to treatments to further preserve muscle integrity and function.

PrgmNr 3788 - Translating precision medicine into the world of speech-language pathology: First follow-up results of the proactive Babble Boot CampÂ© intervention in infants with classic galactosemia

View session detail

Author Block: **B. Peter**[1], J. Davis[2], L. Finestack[3], C. Stoel-Gammon[4], M. VanDam[5], D. Williams[5], N. Scherer[1], N. Potter[5]; [1]Arizona State Univ, Tempe, AZ, [2]Arizona State Univ., Tempe, AZ, [3]Univ. of Minnesota, Tempe, MN, [4]Univ. of Washington, Seattle, WA, [5]Washington State Univ., Spokane, WA
Disclosure Block: **B. Peter:** None.
Precision medicine is an active area of research and practice in medical fields such as cardiology and oncology, where known genotype-phenotype associations drive proactive and personalized management. Because genetic influences on disorders of spoken language are less well understood, principles of precision medicine are less commonly practiced in speech-language pathology. Here, we present preliminary results from the first clinical trial of a proactive treatment focusing on communication abilities in infants with classic galactosemia (CG). Diagnosed via newborn screening, this metabolic disease is caused by a recessively inherited disruption of the GALT gene. Among the most common traits are severe delays in speech sound production and expressive language, seen in ~60% of children with CG. Typically, speech and language skills cannot be measured until child age 2 to 3 years, precluding early identification and treatment. Because newborns with CG are at known predictable risk for these disorders, they are the ideal participants in a clinical trial of a proactive intervention.
Babble Boot CampÂ© (BBC) is the first comprehensive proactive intervention focusing on communication skills, beginning at age Of 12 children with CG who underwent the BBC, all had expressive language skills in the typical range at follow-up, and all but one had typical speech production skills. These findings are above expectation under random conditions (Fisherâ□□s exact p = .014 for speech and .002 for language). Of three untreated controls with CG, one had speech and language scores â□¤ 1st percentile, while two had typical scores in both areas. These findings are consistent with a beneficial effect of proactive behavioral management of symptoms of a genetic disease. A better understanding of the genetic influences on speech and language disorders may motivate more widespread use of proactive and personalized interventions in these disorders.

PrgmNr 3789 - A Clinicopathologic Study of Malignancy in VCP Associated Multisystem Proteinopathy

View session detail

Author Block: **A. Shmara**; UCI, Irvine, CA
Disclosure Block: **A. Shmara:** None.
Valosin containing protein (VCP) is an important protein with many vital functions mostly related to the ubiquitin-proteasome system that provides protein quality control. VCP-associated inclusion body myopathy with Paget disease of bone and frontotemporal dementia (IBMPFD) is an autosomal dominant disorder caused by mutations in the *VCP* gene on human chromosome 9. *VCP* mutations are also associated with amyotrophic lateral sclerosis (ALS), Parkinson's disease and Charcot-Marie-Tooth disease type 2. VCP has also been strongly involved in cancer, with over-activity of VCP found in several cancers such as prostate and pancreatic cancers, endometrial and esophageal carcinomas, and osteosarcoma. The overexpression of VCP has been associated with poor prognosis and increased metastasis, proving it valuable as a marker for the advancement of these cancers. The inhibition of VCP has been suggested as a treatment of metastasis in certain cancers. Since IBMPFD is caused by gain of function mutations in *VCP*, our hypothesis was that we would find an increased tendency of developing malignancies amongst our patients. We present cases of unusual tumors in patients with classic features of *VCP* associated disease including malignant peripheral nerve sheath tumor, anaplastic pleomorphic xanthoastrocytoma with multiple recurrences, thymoma as well as common cancers. These findings expand the phenotype of IBMPFD to potentially include unusual cancers and highlight the importance of *VCP* mutations in cancer development.

PrgmNr 3790 - Childhood and early onset glaucoma classification and genetic profile in a large Australasian disease registry

View session detail

Author Block: **L. Knight**[1], J. B. Ruddle[2,3,4], D. A. Taranath[1], I. Goldberg[5], J. E. H. Smith[5,6,7], G. Gole[8], M. Y. Chiang[9], F. Willett[9], G. Dâ□□Mellow[10], J. Breen[11,12,13], A. Qassim[1], S. Mullany[1], J. E. Elder[2,14], A. L. Vincent[15,16], S. E. Staffieri[2,4], L. S. Kearns[4], D. A. Mackey[17,18], S. Luu[1], O. M. Siggs[1,19], E. Souzeau[20], J. E. Craig[1]; [1]Dept of Ophthalmology, Flinders Univ., Bedford Park, Australia, [2]Dept. of Ophthalmology, Royal Childrenâ□□s Hosp., Melbourne, Australia, [3]Ophthalmology, Dept of Surgery, Univ. of Melbourne, Melbourne, Australia, [4]Ctr. for Eye Res. Australia, Royal Victorian Eye and Ear Hosp., Melbourne, Australia, [5]Discipline of Ophthalmology, Save Sight Inst., Faculty of Med. and Hlth., Univ. of Sydney, Sydney, Australia, [6]Dept. of Ophthalmology, The Children's Hosp. at Westmead, Sydney, Australia, [7]Dept. of Ophthalmology, Macquarie Univ. Hosp., Sydney, Australia, [8]Univ. of Queensland Childrenâ□□s Hlth.Queensland Clinical Unit, Queensland Childrenâ□□s Hosp., Brisbane, Australia, [9]Dept. of Ophthalmology, Queensland Childrenâ□□s Hosp., Brisbane, Australia, [10]Terrace Eye Ctr., Brisbane, Australia, [11]South Australian Genomics Ctr., South Australian Hlth.& Med. Res. Inst., Adelaide, Australia, [12]Robinson Res. Inst., Univ. of Adelaide, Adelaide, Australia, [13]Adelaide Med. Sch., Univ. of Adelaide, Adelaide, Australia, [14]Dept. of Paediatrics, Univ. of Melbourne, Melbourne, Australia, [15]Dept. of Ophthalmology, New Zealand Natl. Eye Ctr., Faculty of Med. and Hlth.Sci., Univ. of Auckland, Auckland, New Zealand, [16]Eye Dept., Greenlane Clinical Ctr., Auckland District Hlth.Board, Auckland, New Zealand, [17]Lions Eye Inst., Nedlands, Australia, [18]Menzies Inst. for Med. Res., Univ. of Tasmania, Hobart, Australia, [19]Garvan Inst. of Med. Res., Darlinghurst, Australia, [20]Flinders Univ., Bedford Park, Australia

Disclosure Block: **L. Knight:** None.

Purpose: To report the relative frequencies of childhood and early-onset glaucoma subtypes and their genetic findings, in a large single cohort.**Design:** Retrospective clinical and molecular study.**Participants:** All individuals with childhood glaucoma (diagnosed 0 to Methods: We retrospectively reviewed the referrals of all individuals with glaucoma diagnosed at Main Outcome Measures**:** The phenotype and genotype distribution of glaucoma diagnosed at Results**:** 290 individuals (533 eyes) with childhood glaucoma and 370 individuals (686 eyes) with early-onset glaucoma were referred to the ANZRAG. Primary glaucoma was the most prevalent condition in both cohorts. In the childhood cohort, 57.6% of individuals (167/290, 303 eyes) had primary congenital glaucoma (PCG) and 19.3% (56/290, 109 eyes) had juvenile open-angle glaucoma (JOAG). JOAG constituted 73.2% of the early-onset glaucoma cohort (271/370, 513 eyes). Genetic testing in probands resulted in a diagnostic yield of 24.7% (125/506) and a reclassification of glaucoma subtype in 10.4% of probands (13/125). The highest molecular diagnostic rate was achieved in probands with glaucoma associated with nonacquired ocular anomalies (56.5%). Biallelic variants in *CYP1B1* (n=29, 23.2%) and heterozygous variants in *MYOC* (n=24,19.2%) and *FOXC1* (n=21, 16.8%) were most commonly reported among probands with a molecular diagnosis. Biallelic *CYP1B1* variants were reported in twice as many female individuals as male individuals with PCG (66.7% vs 33.3%, P = 0.02). **Conclusion:** We report on the largest cohort of individuals with childhood and early-onset glaucoma from Australasia using the CGRN classification. Primary glaucoma was most prevalent. Genetic diagnoses ascertained in 24.7% of probands supported clinical diagnoses and genetic counselling. International collaborative efforts are required to identify further genes as the majority of cases still lack a clear molecular diagnosis.

PrgmNr 3791 - Deciphering the role of PTPN14 in BMP9 signaling and vascular malformations in Hereditary Hemorrhagic Telangiectasia

View session detail

Author Block: **D. T. Beleford**, O. Mamai, S. Basu, R. J. Akhurst; Univ California San Francisco, San Francisco, CA
Disclosure Block: **D.T. Beleford:** None.
Hereditary Hemorrhagic Telangiectasia (HHT) is a rare genetic condition characterized by multiple arteriovenous malformations (AVMs). AVMs may form in different organs including lung (PAVM), brain, liver, and skin. Catastrophic hemorrhage from visceral or cerebral AVMs may occur and are sometimes fatal. Most cases of HHT are caused by heterozygous loss of function variants in *ENG* (HHT1) or *ACVRL1* (aka *ALK1*) (HHT2), which encode endothelial receptors for TGF-Î²s and BMPs. Our group previously showed that single nucleotide polymorphisms (SNPs) in the genetic modifier *PTPN14* (protein tyrosine phosphatase, nonreceptor-type, 14, OMIM 603155) associate with PAVMs in both HHT1 and HHT2. The current study investigates the molecular genetics of interactions between *Eng* and *Ptpn14* during PAVM formation in HHT, in part by characterizing a novel conditional endothelial cell-specific knockout for *Ptpn14*. Our overarching hypothesis is that PTPN14 interacts with BMP9-Smad/Hippo-TAZ signaling pathways to stabilize blood vessels such that reduced expression or loss of *PTPN14* will potentiate the formation of AVMs under conditions of genetic loss of an *ENG* allele. To test this hypothesis, we will investigate the natural development of the retinal vascular network during the first seven days of life in a novel, conditional *Cdh5-Cre*^{ERT2}*.Ptpn14*$^{flox/flox}$ mouse line that, upon treatment with tamoxifen, activates cre recombinase to delete the *Ptpn14* gene specifically in endothelial cells. The novel mouse *Ptpn14*^{flox} allele co-expresses mCherry from the endogenous *Ptpn14* gene reporter before tamoxifen induced Cre-recombinase deletion, and eGFP from the gene promoter after deletion of *Ptpn14.* The characterization of this mouse line will be presented. Primary endothelial and microvascular mouse cell lines will also be developed for *in vitro* assessment of Ptpn14 binding partners, including Smads and Hippo pathway components. Ultimately, this study will impact on our understanding of the pathogenesis of vascular malformations and hemorrhage control in both HHT and other vascular conditions, and may provide novel insights into therapies for HHT.

PrgmNr 3792 - Genotype-phenotype correlations within type IV collagen glomerulopathy

View session detail

Author Block: **M. Elliott**[1], N. Vena[2], M. Marasa[1], A. G. Gharavi[3]; [1]Columbia Univ. Irving Med. Ctr., New York, NY, [2]Hammer Hlth.Sci. Building, New York, NY, [3]Columbia Univ, New York, NY
Disclosure Block: **M. Elliott:** Royalty(ies)/Honoraria; Otsuka Canada.
Type IV collagen glomerulopathy is the second most common cause of genetic kidney disease and is associated with a range of phenotypes. We assessed the phenotypic spectrum associated with type IV collagen variants in a cohort of 2465 kidney disease patients who underwent exome sequencing as part of a biobanking study. Multiethnic patients were recruited at Columbia University. Variants in *COL4A3, COL4A4* and *COL4A5* were interpreted per ACMG guidelines and modified per ClinGen Sequence Variant Interpretation Working Group and Savige *et al*. Patients were categorized as having a pathogenic/likely pathogenic (P/LP) variant or suspicious variant of uncertain significance (VUS). 128 patients with a variant in *COL4A3, COL4A4* or *COL4A5* were identified (5% of cohort) including 3 digenic and 8 compound heterozygote cases. 39% carried a diagnosis of Alport syndrome (AS), 16% focal segmental glomerulosclerosis (FSGS), 25% CKD of unknown cause, 9% thin basement membrane and the remaining 10% carry distinct clinical diagnoses. 75 cases had a kidney biopsy with 34 showing FSGS. Within the 21 cases presenting with a diagnosis of FSGS, 32 with CKD of unknown cause, and 34 with biopsy-proven FSGS, extra-renal symptoms of AS are present in 1, 4, and 6 and a suggestive family history was present in 4, 13, and 9 cases respectively. We identified 137 variants including 21, 34 and 52 P/LP variants and 10, 8 and 12 VUS in *COL4A3*, *COL4A4*, and *COL4A5*, respectively. Of the 108 P/LP variants: 75 missense, 3 nonsense, 16 frameshift, 14 splice-site. VUS covered all variant types except frameshift and truncating. Clinical features did not differ between cases with variants classified as P/LP or VUS. Cases with variants in *COL4A5* reached kidney failure at an earlier age than *COL4A3* or *COL4A4* (42 vs 65 vs 58 years, p=0.0003), but no differences were noted in hematuria, proteinuria, kidney transplantation and failure rates, hearing loss, ocular abnormalities, or a family history of the above. Digenic and compound heterozygote cases had similar rates of kidney failure as single variant cases (64% vs 47%, p=0.55). Frameshift variants associate with hearing loss, receiving a kidney transplant and developing kidney failure at a younger age compared to the other variant types, but other clinical features including kidney failure rate were similar. Clinical phenotype was similar across *COL4A3,4,5* genes. Frameshift variants associate with hearing loss and age of kidney failure. Many patients with an identified variant in *COL4A3, COL4A4* or *COL4A5* present with FSGS or CKD of unknown cause and lack extra-renal symptoms of AS or a concerning family history, showing the importance of genetic testing in these populations.

PrgmNr 3793 - Lower penetrance estimates for neurosusceptibility loci using an improved formula

View session detail

Author Block: **S. Goh**[1], R. Bowden[2], M. Pinese[3], E. Kirk[1]; [1]Sydney Children's Hosp., Sydney, Australia, [2]Monash Univ., Melbourne, Australia, [3]Unviersity of New South Wales, Sydney, Australia
Disclosure Block: **S. Goh:** None.
Introduction: Accurate penetrance estimates are important for diagnosing genetic conditions and providing prenatal advice. Penetrance has been defined as the proportion of individuals with a given genetic change who display a phenotypic change. However, we show that this definition is ambiguous and has led to a misunderstanding when applied to a mathematical formulation of Bayesâ theorem. As a result, penetrance estimates of many neurosusceptibility loci have been overestimated.
Methods: We provide a mathematical rationale for a more clinically-meaningful formula for estimating penetrance. We also identify that the background rate of intellectual disability is an important variable for the formula. Therefore, we use clinically-appropriate cohorts of individuals who have intellectual disability, rather than any disability, in our calculations.
Results: We show that penetrance estimates for most neurosusceptibility loci are markedly lower than previously published. Small 15q13.3 (*CHRNA7*) duplications are shown to have a penetrance of 0%. Others like 16p11.2 (*SH2B1*) distal duplications, 16p13.11 (*MYH11*) duplications, large 15q13.3 (*CHRNA7*) duplications and 17p13.3 (*YWHAE*) duplications, have a recalculated penetrance for intellectual disability of 1-5%, with a 95% confidence interval that may include 0%. Some higher penetrance estimates of 40% show a recalculated penetrance of approximately 10-15%.
Conclusion: Most neurosusceptibility loci have a lower penetrance than previously estimated. Many higher published penetrance estimates in previous works were based on a mathematical misunderstanding in the formulation of Bayesâ Theorem. This finding corrects a long-standing iatrogenic miscalculation of penetrance estimates for copy number variants and has significant diagnostic and counselling implications in both the prenatal and postnatal settings.

PrgmNr 3795 - Analytical performance of germline genome sequencing for clinical exome diagnostic testing

View session detail

Author Block: **N. Hammond**[1], P. Tong[1], D. Fisk[1], S. White[1], E. Spiteri[1,2], E. Ashley[3], S. A. Scott[1,2], Y. Yang[1,2]; [1]Clinical Genomics Lab., Stanford Hlth.Care, Palo Alto, CA, [2]Dept. of Pathology, Stanford Univ., Palo Alto, CA, [3]Stanford Ctr. for Inherited Cardiovascular Disease, Palo Alto, CA
Disclosure Block: **N. Hammond:** None.
Diagnostic panel and exome testing commonly employ target enrichment-based sequencing; however, germline genome sequencing has potential technical advantages, including relatively unbiased coverage, copy number and structural variant detection, and flexible interrogation of clinical regions of interest (ROI). In addition to diagnostic testing, genome sequencing will likely have potential utility for carrier screening, pharmacogenomic and polygenic risk score screening, and molecular karyotyping, among other applications. Given the potential utility of genome sequencing for clinical exome testing, we evaluated its analytical performance compared to enrichment-based exome sequencing. A commercial exome ROI was intersected with our novel clinical exome ROI, and the shared genomic region was utilized for performance evaluation across Genome in a Bottle (GIAB) reference material samples (n=5) that were subjected to both enrichment-based and genome sequencing with small variant (98% exome ROI coverage, which required a depth of >140X using enrichment-based sequencing. Compared to high-confidence GIAB truth sets, overall single nucleotide variant (SNV) recall and precision within the exome ROI were >99.8% for enrichment-based and genome sequencing. However, the accuracy for insertion/deletion (indel; 1-50 bp) detection was markedly enhanced by genome sequencing, as indel recall and precision within the exome ROI were 96.8% and 95.5% for enrichment-based sequencing, compared to 99.7% and 99.7% for genome sequencing, respectively. Moreover, genome stratification determined that SNV/indel accuracy was notably improved across GIAB-defined difficult regions that intersected the exome ROI, with recall and precision of 94.9% and 93.1% for enrichment-based sequencing, compared to 99.6% and 99.6% for genome sequencing. Genome sequencing across the exome ROI was also very robust, as the average SNV/indel inter- and intra-run, inter-instrument, and inter-flowcell non-reference genotype concordance was 98.5%. These results indicate that clinical diagnostic exome testing using genome sequencing has superior small variant calling compared to enrichment-based sequencing and strongly support the implementation of this technology when cost-effectiveness can be justified by clinical laboratories.

PrgmNr 3796 - Clinical validation of germline genome sequencing copy number variant confirmation using qPCR

View session detail

Author Block: **C. Reavey**[1], C. Ho[1], M. E. Grove[1], L. Villaneuva[1], L. Chen[1], K. Dumas[1], K. S. Weymouth[1], P. Tong[1], L. Gojenola[1], E. Spiteri[1,2], Y. Yang[1,2], S. A. Scott[1,2], W. Qiao[1,2]; [1]Clinical Genomics Lab., Stanford Hlth.Care, Palo Alto, CA, [2]Dept. of Pathology, Stanford Univ., Stanford, CA
Disclosure Block: **C. Reavey:** None.
Copy number variants (CNVs) are increasingly implicated in Mendelian genetic disorders and as risk alleles for complex diseases with incomplete penetrance, prompting their recent incorporation into diagnostic panel testing. In addition to microarrays and other hybridization-based technologies, germline CNVs can be inferred from short-read multi-gene panels, exome and genome sequencing by several publicly available bioinformatic algorithms. However, the accuracy and precision of sequencing-based CNV detection necessitates orthogonal confirmation for most clinical testing applications. To facilitate the implementation of genome sequencing-based diagnostic panel testing and CNV calling using Illumina DRAGEN pipeline (Manta and CNV Caller), we developed and validated a rapid and flexible quantitative polymerase chain reaction (qPCR) assay to confirm clinically significant germline CNVs identified by clinical genome sequencing. A total of 63 CNV control specimens and publicly available reference material samples characterized by orthogonal methods were included in this qPCR validation. The evaluated performance characteristics included accuracy, precision (repeatability and reproducibility), efficiency and linear range, reference interval (copy number range), analytical sensitivity, and analytical specificity. The qPCR assay employed the KAPA SYBR FAST qPCR kit with amplification performed on the LightCycler 480 instrument II, and data analysis using the 2-ddCt method. PCR efficiency and standard curves were evaluated (average R^2 >99.7), which enabled the definition of ratio ranges for germline copy numbers ranging from zero to four. Importantly, assessment of all CNV control specimens (deletions, duplications, triplications) by qPCR resulted in an analytical sensitivity, specificity, and accuracy of >99.9% across all interrogated autosomal and sex chromosome copy number values (0-4). In addition, the germline qPCR assay was very robust, as repeatability and reproducibility of copy number calling were completely concordant across triplicate testing (relative ratio standard deviation

PrgmNr 3797 - Detecting absence of homozygosity using high-resolution optical genome mapping

View session detail

Author Block: **A. R. Rao**, E. T. Lam, J. A. Velazquez-Muriel, D. Zhang, N. Miller, A. W. C. Pang, A. R. Hastie, A. Chaubey, M. Oldakowski; Bionano Genomics, San Diego, CA
Disclosure Block: **A.R. Rao:** Salary/Employment; Bionano Genomics.
Copy-neutral loss of heterozygosity (CN-LOH), which could be a result of uniparental disomy, is associated with meiotic errors resulting in developmental diseases. CN-LOH may also occur as a somatic event in several cancers. Events have traditionally been identified using microarray and microsatellite analysis. Optical genome mapping (OGM) has the potential to replace most of these traditional analyses. Using information about structural variant (SV) call zygosity, it is also possible to infer the absence or loss of heterozygosity (AOH/LOH). Here we describe a method for AOH analysis based on OGM results from the Bionano Genomics Saphyr platform. Regions of homozygosity are identified by a consistent decrease in heterozygous SV calls across a genomic region in the case sample compared to the level observed genome-wide in controls. DNA was processed on the Bionano Genomics Saphyr instrument and analyzed using the Bionano Solve pipeline. After filtering SV calls for high-quality, informative sites, AOH events were simulated by splicing together SV calling datasets from 153 controls and 4 haploid samples, where regions derived from haploid genomes represented AOH events. A Hidden Markov Model (HMM) was used to model the spatial dependence between neighboring SVs of a given zygosity, and the model parameters were estimated by fitting the model to the simulated dataset. Performance was evaluated on newly simulated samples and additional samples with known AOH/LOH events that were previously identified using microarray technology. In simulated data, our method achieved high sensitivity and precision in detecting large AOH regions typical of the sizes that underpin uniparental disomy, with 92% sensitivity and 97% precision, where true positives were defined as an overlap of 80% between simulated and predicted regions. Using a small cohort of samples with known AOH/LOH events, we have validated the utility of this method, and were able to detect high variant allele frequency events larger than 10 Mbp. Our results show that it is possible to detect AOH/LOH regions using OGM alone. Future developments incorporating additional data will greatly improve the resolution to reliably detect events of smaller size, and those with low variant allele frequency, which would expand the utility of the method to analyzing tumor samples. The software is available as part of the Bionano Access v1.7.

PrgmNr 3798 - Detection of *RHD*/*RHCE* hybrid genes using the Axiomâ¢ Universal Blood Donor Typing research array

View session detail

Author Block: **R. Varma**[1], M. Aguad[2], A. Roter[1], J. Schmidt[1], W. J. Lane[2], J. Gollub[1], on behalf of the Blood transfusion Genomics Consortium; [1]Thermo Fisher Scientific, Santa Clara, CA, [2]Brigham and Womenâ€™s Hosp. / Harvard Med. Sch., Boston, MA
Disclosure Block: **R. Varma:** Salary/Employment; Thermo Fisher Scientific.
Blood transfusion requires matching blood types between donor and recipient to avoid hemolytic reactions. The most important blood group systems are ABO and Rh. The Rh system includes five main antigens (D, C, c, E and e) which are encoded by the RHD and RHCE genes. These genes are highly homologous and inversely adjacent on chromosome 1p36.11. Hybrid RH alleles have evolved by replacements of RHD fragments with corresponding regions from the RHCE gene or vice versa. Detection of such hybrid genes is critical for accurate Rh antigen typing. The Applied Biosystemsâ¢ Axiomâ¢ Universal Blood Donor Typing (UBDT_PC1) research array has been designed to detect variants underlying erythrocyte, leukocyte, platelet, and neutrophil antigens. Detection of structural variants within RHD and RHCE is enabled using about 500 probesets interrogating mostly intronic sequences over the entire length of both genes. Measured probeset signal intensities are used to estimate copy number in defined regions. Copy number changes within each gene provide evidence of rearrangements and thereby enable hybrid gene detection. Cell line genomic DNA samples (N=270) from four HapMap populations were assayed on the UBDT_PC1 array. Several rearrangement events were detected with high reproducibility on the RHD and RHCE genes and were compared to high coverage Whole Genome Sequencing (WGS) data of 1000 Genomes Project (1000G) samples. Seven YRI (Yoruba in Ibadan, Nigeria) samples, including members of two families, showed copy number decreases over RHD exons 4 to 7 with normal diploid copy number states on flanking exons. In these same samples, copy number increases were detected in RHCE exons 4 to 7, indicating potential rearrangement events. These samples were confirmed as heterozygous or homozygous DIIIa-CEVS(4-7)-D by WGS, which is a D- allele (RHD*03N.01) found in African populations. One CHB (Han Chinese in Beijing, China) sample showed a copy number decrease over RHD exons 3 to 9 with a corresponding increase in RHCE. High coverage WGS data identified this individual as a D-CE(3-9)-D heterozygote type, which is a D- allele (RHD*01N.04) found in Asian populations. One sample with European ancestry showed a copy number decrease over RHCE exons 2 to 5 with a corresponding increase in RHD. This individual was identified as a novel CE-D(2-5)-D heterozygote by WGS. Other events indicative of the common C+ antigen RHCE*Ce or *CE allele have also been detected on the UBDT_PC1 array and confirmed by WGS. These results show that the Axiom array platform accurately detects complex structural variants of the RH genes and offers a promising approach for analyzing blood types in transfusion research applications.

PrgmNr 3799 - Identification of structural variants in constitutional disorders by optical genome mapping

View session detail

Author Block: **A. Pang**[1], H. Barseghyan[1,2], A. Hastie[3], A. Chaubey[1]; [1]Bionano Genomics, San Diego, CA, [2]Children's Natl. Res. Inst., Washington, DC, DC, [3]BioNano Genomics, San Diego, CA
Disclosure Block: **A. Pang:** Salary/Employment; Bionano Genomics.
Chromosomal structural variation contributes towards most constitutional genetic disorders. Karyotyping, fluorescence in situ hybridization (FISH) and chromosomal microarrays (CMA) are currently used in clinical cytogenetics and molecular diagnostics. CMA is the first line test recommended by the American College of Medical Genetics (ACMG), American Academy of Pediatrics (AAP), and America Academy of Neurology (AAN); however, the limitations of CMA include the inability to detect balanced translocations and inversions, and it cannot differentiate segmental duplication and repetitive regions of the genome.
Recent advances in chromosomal analysis with optical genome mapping (OGM) have the potential to address these shortcomings. Bionano Genomicsâ Saphyr platform images linearized, megabase-size, DNA molecules, labeled at specific sequence motifs in nanochannel arrays. Imaged molecules (>150kbp) are de novo assembled or aligned directly to the human genome reference for structural variant (SV) calling. OGM can detect germline SVs >500 bp in size, including repeat expansions (FMR1) and contractions (DUX4), balanced events, such as inversions, reciprocal translocations, and more recently, copy-neutral regions with absence of heterozygosity (AOH). SV population frequencies are estimated by comparison against a control database consisting of OGM samples from >26 populations with no reported disease phenotype.
OGM was applied to study the genomes of individuals with multiple constitutional disorders. Segmental duplication mediated deletions such as a 1.9 Mbp deletion in 7q11.23 (Williams-Beuren Syndrome), and a 3.7 Mbp deletion in 17p11.2 in (Smith-Magenis Syndrome). OGM also captured autosomal trisomies (+13 and +21), monosomy X, t(2;10) and t(9;21) translocations were all detected. In addition to SV detection, the chromosome structures of SVs, such as the location and orientation of duplications, can be elucidated by OGM, and these structures have been validated by FISH analysis.
In conclusion, OGM has the potential to identify a large range of SVs and to improve the characterization detected by conventional methodologies.

PrgmNr 3800 - Predicting Genes from Phenotypes using Human Phenotype Ontology (HPO) terms

View session detail

Author Block: **A. M. Slavotinek**[1], H. Prasad[1], T. Yip[1], S. Rego[1], P-M. Martin[1], J. van Ziffle[1], W. Devine[1], U. Hodoglugil[2], H. Hoban[1], M. Kvale[3]; [1]Univ California San Francisco, San Francisco, CA, [2]Genomic Med. Lab., UCSF, San Francisco, CA, [3]Inst. for Human Genetics, San Francisco, CA
Disclosure Block: **A.M. Slavotinek:** None.
The interpretation of genomic variants following exome sequencing (ES) can be aided by using HPO terms to standardize clinical features and predict gene relevance. The Pediatric arm of the P3EGS study at UCSF has performed ES on 474 probands and parents, largely from underserved populations. We studied patients who received pathogenic (P) or likely pathogenic (LP) variant results and ascertained 114 P or LP variants and corresponding genes. We used PhenoDB to manually extract HPO terms from a single clinical note and used Phen2Gene and the HPO terms to predict and rank the gene corresponding to each causative variant. Phen2Gene gene rankings were assigned to 6 rank classes, with class 1 covering ranks 1-10 and class 2 covering ranks 11-50. We found that the rank class was inversely correlated to the Phen2Gene gene score (p -16). Phen2Gene was able to rank the causative gene into rank classes 1 or 2 in just over 25% of cases. Most of the genes in rank class 1 were associated with well-characterized phenotypes, such as Noonan syndrome. We then examined the effects of patient and gene variables to affect the Phen2Gene scores and rank classes. A linear regression analysis showed a significant association between the number of years since the gene was first published with higher Phen2Gene scores (p -16; Bonferroni threshold p 0.5 was reported 9 or more years ago. There was a significant correlation between a higher number of OMIM terms associated with the gene and a higher score (p = 0.0011) and this relationship demonstrated a threshold effect, as only one gene with 0.5). The number of terms with an HPO hierarchal depth greater or equal to 11 was also statistically significant (p = $2.5x10^{-4}$), implying that terms deep in the HPO hierarchy had the best chance of producing a high scoring gene. Autosomal dominant inheritance of the gene and a diagnostic category of â□□ID with MCAâ□□ also trended towards a relationship with score but were not significant. Many causative genes were not highly ranked, possibly because of non-specific presentations, rarity of the gene, and inclusion of HPO terms distracting to the phenotype. We conclude that gene prediction using HPO terms and software tools may play a role in association with well delineated and distinctive clinical presentations, especially for established genes. The significant associations with Phen2Gene score for a higher number of OMIM terms and number of terms with HPO depth greater or equal to 11 suggests that pleiotropy, comprehensive phenotyping, and specificity of clinical terms may be important factors in successful gene prediction.

PrgmNr 3801 - Resolving complex pathogenic alleles using HiFi sequencing for long-range amplicon data with a new clustering algorithm

View session detail

Author Block: **J. Harting**[1], C. Heiner[2,1], I. McLaughlin[3], Z. Kronenberg[1]; [1]Pacific BioSci.s, Menlo Park, CA, [2]LA HONDA, CA, [3]Pacific BioSci.s of California, Inc., Menlo Park, CA
Disclosure Block: **J. Harting:** Salary/Employment; Pacific Biosciences.
Many genetic diseases are mapped to structurally complex loci. These regions contain highly similar paralogous alleles (>99% identity) that span kilobases within the human genome. Comprehensive screening for pathogenic variants amongst paralogous sequences is incomplete and labor intensive using short reads or optical mapping. In contrast, long-range targeted amplification and PacBio HiFi sequencing fully and directly resolve and phase a wide range of pathogenic variants without assembly or inference. To capitalize on the accuracy of HiFi data, we designed a new amplicon analysis tool, pbaa, which uses a newly developed sequence clustering algorithm to rapidly deconvolve (separate) a mixture of haplotypes, enabling precise diplotyping, and disease allele classification.
In this experiment, we analyzed two sets of gene-pseudogene systems, GBA and CYP, that are the second and eighth most common carrier disease alleles, respectively. Samples tested were selected from the Coriell catalog known to have pathogenic variants troublesome to test for with standard short-read assays. Co-amplified long-range PCR amplicons were generated for GBA (12 kb)/GBAP1 (15 kb) responsible for Gaucher disease, as well as CYP21A2 (10 kb)/CYP21A1P (8 kb) responsible for congenital adrenal hyperplasia. We obtained 7 samples to test the CYP21A2 region and 13 separate samples for GBA. HiFi reads were then generated from the amplicon libraries on both Sequel and Sequel II Systems, with replicated samples, to achieve a 24-sample multiplex for each target. Consensus amplicons were produced using pbAA and variants were determined using minimap2 alignments along with a custom SQL database for characterizing and reporting results. From these data we were able to accurately call all pathogenic variants in the test samples for all replicates, including whole-gene deletions, gene duplication, gene fusions, recombinant exons, and phased complex heterozygotes. In one trio affected by adrenal hyperplasia, three large structural variants were correctly and independently attributed to the parents and proband, including a duplication of CP21A1P and a CYP21A1P-CYP21A2 gene fusion in the mother and a CYP21A2 deletion in the father. This experiment demonstrates how PacBio HiFi data, analyzed with pbaa, simplifies targeted disease allele identification.

PrgmNr 3803 - Biallelic loss of function ATM due to pathogenic synonymous and novel deep intronic variant c.1803-270T>G identified by genome sequencing in child with ataxia-telangiectasia

View session detail

Author Block: **T. Maroilley**[1,2,3], N. A. M. Wright[3,4], C. Diao[1,2,3], L. MacLaren[2,3], G. Pfeffer[2,5], J. R. Sarna[5], P. Y. B. Au[2,3,6], M. Tarailo-Graovac[1,2,3,6]; [1]Dept. of Biochemistry and Molecular Biology, Cumming Sch. of Med., Univ. of Calgary, Calgary, AB, Canada, [2]Dept. of Med. Genetics, Cumming Sch. of Med., Univ. of Calgary, Calgary, AB, Canada, [3]Alberta Children's Hosp. Res. Inst., Univ. of Calgary, Calgary, AB, Canada, [4]Section of Pediatric Hematology-Immunology, Dept. of Pediatrics, Alberta Children's Hosp., Univ. of Calgary, Calgary, AB, Canada, [5]Dept. of Clinical NeuroSci.s and Hotchkiss Brain Inst., Univ. of Calgary, Calgary, AB, Canada, [6]contributed equally, Calgary, AB, Canada
Disclosure Block: **T. Maroilley:** None.
The majority of rare disease (RD) patients remain undiagnosed or partially diagnosed. Genome sequencing has the potential to explore non-coding regions and help increase the diagnosis. However, interpreting the pathogenicity of variants in non-coding regions is challenging. Ataxia-telangiectasia (AT) is a complex neurodegenerative RD characterized by progressive ataxia and movement disorder (chorea and dystonia), telangiectasias, immune defects and increased risk of malignancy. AT is caused by biallelic loss of function variants in *ATM,* which encodes a phosphatidylinositol 3-kinase that responds to DNA damage by regulating DNA repair and cell cycle control pathways. Here we report a child with progressive ataxia and chorea. A GeneDx clinical ataxia gene panel identified maternally inherited synonymous variant (NM_000051.3: c.2250G>A: p.Lys750=) but no second variant. The c.2250G>A variant was previously described to result in exon 14 skipping. Although AFP in this child was normal, chromosome analysis identified the presence of frequent t(7;14) and t(7;22) translocations, which was highly suggestive of AT. To identify the second variant in *ATM*, we performed genome sequencing. We discovered a deep intronic variant (NM_000051.3: c.1803-270T>G) inherited from father. Transcript analyses revealed that c.1803-270T>G variant results in aberrant splicing and the inclusion of 56 base-pairs of intron 11. *In silico* tests predict that this causes a premature stop codon, suggesting likely non-functional ATM. Exon 14 skipping due to the maternal synonymous variant was also confirmed. Functional analyses by DNA damage repair assessment in lymphocytes in blood by flow cytometry (DDRFL; Nationwide Children's) revealed substantial decrease in pATM, pSMC1 and gamma H2AX in T, B and NK cells at 1h post-irradiation with 2Gy. Furthermore, persistence of gamma H2AX 24hr post-radiation was indicative of impaired DNA double-strand break repair, consistent with abnormal pATM function. Together, these analyses confirmed biallelic functional loss of ATM. The unanticipated finding of compound heterozygosity for a synonymous and a deep intronic variant further highlights the power of genome sequencing in reducing missing heritability by enabling consideration of more complex genetic mechanisms when diagnosing RD patients. Importantly, even though a clinical diagnosis of AT was highly suspected here, identification of the second variant not only confirmed the diagnosis for the child but also has significant clinical implications for the detection of carriers, who are at risk for cancer predisposition and require preventative surveillance and management.

PrgmNr 3804 - End ALS Kaggle challenge: Insights from aberrant transcriptome analysis

View session detail

Author Block: **M. Celik**[1], F. Brechtmann[2], J. Gagneur[3], F. HÃ¶lzlwimmer[4], M. MÃ¼eller[4], X. Xie[1], V. YÃ©pez[4], M. Zech[5], N. Wagner[4]; [1]UC Irvine, Irvine, CA, [2]Technical Univ. Munich, Munich, Germany, [3]Technical Univ. of Munich, Garching, Germany, [4]Technical Univ. of Munich, Munich, Germany, [5]Munich, Germany

Disclosure Block: **M. Celik:** None.

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative neuromuscular disease with poor survival prognosis and no cure. To progress in our understanding of ALS, Answer ALS, EverythingALS and Roche Canada's Artificial Intelligence Centre of Excellence initiated an AI modeling challenge via the platform Kaggle providing participants with genetic, transcriptome and phenotypic data of 138 ALS patients and 32 controls. We present here the winning analysis to the challenge task 1 (One mechanism of action or multiple independent mechanisms of action?).

Given the heterogeneity of genetic causes of ALS, we assumed that no single gene or pathway would collectively explain all cases. Therefore, we developed an approach focusing on discovering aberrations that could be specific for a single or a few cases. To this end, we first identified expression outliers using OUTRIDER, a denoising autoencoder that we previously adapted to work with RNA-seq data. To ensure that the outliers are driven by genetics, we filtered them down to genes that carried rare genetic variants likely disrupting gene expression (VEP high and moderate impact, MMSplice and SpliceAI). This analysis yielded 109 genetically-supported outlier genes.

Among these, three genes (NEK1, OPTN, SPG11) corresponded to catalogued ALS mutations or known ALS genes, supporting the validity of the approach and possibly yielding a more detailed genetic diagnosis of those patients. We then searched for new ALS genes using a two-fold strategy. First, we considered genes functionally linked to established ALS genes by applying network diffusion on the STRING gene network. This revealed 16 genetically-supported outliers which are functionally linked to ALS genes or are well established disease genes. Second, we asked whether new potential pathways would emerge as clusters of functionally related genetically-supported outliers. This second analysis revealed gene groups that are members of nucleopores, kinetochores, DNA repair and machinery, as well as ribosomal RNA biogenesis.

In conclusion, our main results include:

- variants associated with aberrant expression for known ALS genes, potentially characterising those affected patients
- new high impact variants in further cases in a gene potentially related to ALS, which would improve our catalogue of pathogenic variants
- new candidate genes in known pathways
- new candidate pathways
- methodological showcase of combining expression outlier and network diffusion analysis to genetic predisposition analysis

Our submission and notebook is accessible at tinyurl.com/xh9jr8kt.

PrgmNr 3805 - Establishing a platform for the functional study of oligogenic novel ALS genes

View session detail

Author Block: **S. Wu**[1], J. A. Fifita[2], S. Yang[1], I. P. Blair[3]; [1]Macquarie Univ. Ctr. for Motor Neuron Disease Res., Sydney, Australia, [2]Macquarie Univ., Sydney, Australia, [3]Macquarie Univ, Sydney, Australia

Disclosure Block: **S. Wu:** None.

Amyotrophic lateral sclerosis is a deadly neurodegenerative disease with a genetic basisâ□□the disease is monogenic in 60% of ALS families. The remaining unsolved ALS families do not exhibit classical Mendelian inheritance patterns and are thus challenging to solve via traditional genetic analyses alone. We have developed an *in vitro* functional pipeline using cells transiently expressing each candidate gene variant to prioritize individual candidates based on their capacity to induce known ALS cellular pathology. This pipeline was applied to five candidate gene variants in an ALS family negative for known ALS-causative gene mutations. Of the five, two candidates exhibited strong potential for ALS pathogenicity. Interestingly, the genomic proximity of these two candidate genes suggest that they are likely co-inherited, and potentially produce oligogenic disease effects. This refers to when two or more gene mutations are required for pathogenicity. Although oligogenicity has been implicated in ALS, rapid reliable functional screens are sorely lacking. To investigate the compound effect of both candidates, we attempted to apply our *in vitro* pipeline to cells transiently co-expressing both candidate variants. Unfortunately, low co-transfection and -nucleofection efficiencies necessitated a strategy change. Thus, we sought to express both candidates from a single vector. To accommodate the large size of one candidate gene, we designed a multicistronic vector using two 2A peptide sequences derived from the *porcine teschovirus-1* and *thosea asigna* viruses, which allows candidate gene expression at an equal ratio. Preliminary results showed that co-transfection efficiency dramatically increased and that the spatial-temporal expression pattern of each candidate gene remained consistent with previous findings. We have successfully prioritized two novel candidate ALS genes, and developed a promising molecular platform for the functional study of oligogenicity in ALS that allows the continued application of our *in vitro* functional pipeline in our novel gene discovery strategy. These genes will be further investigated first in zebrafish to provide *in vivo* support for their roles in ALS before progressing to mice models. Disease gene discoveries not only drive our understanding of the mechanisms underlying ALS but also offer fresh diagnostic and therapeutic targets.

PrgmNr 3806 - High-throughput classification of pathogenic *SLC6A1* variants

View session detail

Author Block: **M. Trinidad**, S. Froelich, G. Berguig, W. Wallace, J. H. LeBowitz, K. Estrada, A. Wuster; BioMarin Pharmaceutical, San Rafael, CA
Disclosure Block: **M. Trinidad:** Salary/Employment; BioMarin Pharmaceutical.
Protein-coding variants in *SLC6A1* are associated with early onset epilepsy, autism and schizophrenia. *SLC6A1* encodes the GABA transporter GAT1 and is expressed at the synapses of GABAergic neurons. The primary function of neuronal GAT1 is to re-uptake the inhibitory neurotransmitter GABA from the synaptic cleft in order to sustain its synaptic concentrations. Recently, variants in *SLC6A1* have been associated with a spectrum of epilepsy syndromes and neurodevelopmental disorders.
Pediatric patients present with seizures, neurodevelopmental delay, autism-like behaviors and other neurological and developmental symptoms. The average age of onset of seizures is 3.7 years. *De novo* mutations in *SLC6A1* have also been associated with schizophrenia. Schizophrenia patients with *SLC6A1* mutations are typically diagnosed in early adulthood and do not present with the symptoms seen in pediatric patients.
Although advances in genomic sequencing have enabled aggregation of patient variants and observed clinical manifestations, the functional impact of many of the disease associated variants remains unknown. Experimental characterization remains a valuable method for validating the impact of variants. Here, we present a high-throughput cellular assay to determine the functional impact of more than 180 *SLC6A1* missense-variants on GABA transport. Of those, 88 are seen in epilepsy or neurodevelopmental delay patients and 13 in schizophrenia patients. The remaining mutations are rare variants of unknown significance (VUS).
To expand the understanding of the genotypic and phenotypic spectrum of the disorder, we quantified GABA uptake activity using plasmid expression constructs for each missense variant with HEK293T GAT1-knockout cells. Cells were incubated with deuterated GABA, then analyzed by mass spectrometry to assess GAT1 transporter activity. While the screen is ongoing, our preliminary results find this method to be highly concordant with variant activities previously reported in the literature. These results will allow us to classify variants based on activity to generate a more accurate estimate of the prevalence of *SLC6A1*-associated disorders, and to better understand the molecular biology of *SLC6A1*-associated conditions. The acquired knowledge will help guide therapeutic decisions and enable the development of targeted therapies that enhance transporter function and improve symptoms.

PrgmNr 3807 - Immune cell specific epigenetic and regulatory mapping of variants associated with multiple sclerosis in hispanic populations

View session detail

Author Block: **S. Cruz-Gonzalez**[1], H. Bai[2], A. H. Beecham[3], K. Miskimen[2], C. Manrique[3], W. F. Hulme[4], A. J. Griswold[3], J. McCauley[3], W. S. Bush[5]; [1]Univ. of California, San Francisco, San Francisco, CA, [2]Case Western Reserve Univ., Cleveland, OH, [3]Univ. of Miami, Miami, FL, [4]Univ. of Miami Miller Sch. of Med., Miami, FL, [5]Case Western Reserve Univ, Cleveland, OH
Disclosure Block: **S. Cruz-Gonzalez:** None.
Multiple Sclerosis (MS) is a chronic, autoimmune disease that affects the central nervous system (CNS). MS affects individuals of all ages and often involves periodic â□□attacksâ□□ that severely damage the CNS causing permanent damage. The disease affects approximately 2-3 million individuals worldwide, with prevalence varying by a countryâ□□s economic status. Notably, the prevalence of MS among Latin American individuals has been reported to be increasing in the last decade. Thus, it is essential to generate actionable molecular insights about this disease that can benefit these individuals. Presently, mechanistic understanding for MS is lacking, but previous studies have found MS has a genetic component that includes the major histocompatibility complex (MHC) and 200 single nucleotide polymorphisms (SNPs) outside the MHC. These associated MS variants were discovered almost exclusively in individuals of European ancestry. Functionally, these SNPs have been found to induce regulatory changes to immune cells (e.g.: T cells and Myeloid cells) which have also been associated with the development of MS. While these genetic aspects of MS are well-studied in patients of European descent, their influence on Latin American individuals has not been examined. As a demonstration of approaches that can be used to fill this research gap, we have recruited 9 Puerto Rican MS patients for functional genomics assays. Both whole blood samples and sorted CD4+ T-cells, CD8+ T-cells, and CD33+ monocytes were processed to produce three types of data: genotypes, whole-blood, and sorted-cell RNA-seq, and sorted-cell ATAC-seq. We found that 57 of the 200 MS risk SNPs were present in the Puerto Rican samples (MAF > 1%), with a mean Odds Ratio (OR) of 1.14. We next investigated if any of these SNPs overlapped chromatin accessibility peaks that were significantly associated with an increase in local gene expression (p TNFRSF9 (peak = 6330140-6343170, p = 0.00303), *PPP4C* (peak = chr16: 30085590- 30098400, p = 0.04236), and *CDC14A* (peak = chr1: 100943510- 100955300, p = 0.04198). In conclusion, our functional genomics approach has provided a pioneering overview of the molecular landscape of MS in Puerto Rican individuals, which is a proof of concept for future multi-omics studies of diseases with genetic components in admixed individuals.

PrgmNr 3808 - Paternal allele of origin, radiosensitivity and the effect of genetic modifiers on the 3q29 Microdeletion Syndrome phenotype

View session detail

Author Block: **N. Kopp**[1], J. C. Hodge[2], A. D. Besterman[3], A. Eskin[1], K. M. Squire[1], M. J. Dasouki[4], N. Di Donato[5], R. A. Gatti[6], S. F. Nelson[7], J. A. Martinez-Agosto[8], F. Quintero-Rivera[9]; [1]UCLA, Los Angeles, CA, [2]Indiana Univ., Indianapolis, IN, [3]Rady Children's Hosp. / Univ. of California San Diego (UCSD), San Diego, CA, [4]AdventHlth.Med. Group Genomics and Personalized Hlth., Orlando, FL, [5]Faculty of Med., TU Dresden, Dresden, Saxony, Germany, [6]UCLA Sch. of Med., Los Angeles, CA, [7]UCLA Med Ctr, Los Angeles, CA, [8]Univ California Los Angeles, Los Angeles, CA, [9]Sch. of Med., Univ. of California Irvine (UCI), Orange, CA

Disclosure Block: **N. Kopp:** None.

The 3q29 microdeletion syndrome is a heterogeneous disorder. The cause, a recurrent heterozygous genomic loss spanning approximately 1.6 Mb, likely arises through non-allelic homologous recombination mediated by multiple sets of flanking low-copy repeats. Even within patients of the same deletion size, the phenotype is quite variable, with features that are not specific or clinically diagnostic. The association of psychiatric phenotypes with the 3q29 microdeletion syndrome was first reported in 2010 by Quintero-Rivera and Martinez-Agosto (PMID: 20830797), based on 26 published cases. Since then, more patients have been described with complex psychiatric phenotypes, including autism, bipolar disorder, schizophrenia, attention deficit hyperactivity disorder and depression. Previous studies have reported variants in the *DGL1, SLC51A, and TCTEX1D2* as potential modifiers. Furthermore, recessive loss-of-function mutations in *RNF168*, another gene located in the deleted region, are associated with RIDDLE, a radiosensitivity, syndrome. We performed trio exome sequencing of six patients (3M/3F/Caucasian/Latino/Asian) with 3q29 microdeletion in an effort to identify alterations in their genome, specifically on the remaining allele suggesting a recessive mechanism potentially associated with psychiatric phenotypes. The parent of origin of the deleted allele was determined. Radiosensitivity testing (Colony Survival Assay and Western blot) was also performed on the probandâ□□s lymphoblastoid cell lines to determine the impact of hemizygosity for the *RNF168* gene. We demonstrated that hemizygous disruption of *RNF168* does not affect radiosensitivity. Exome sequencing revealed a range of 35-160 variants in the remaining undeleted 3q29 locus . No additional rare pathogenic variant in a known disease gene was identified. We summarize the 113 variants detected in the deleted region that overlap all cases in this study. Of the variants previously reported, two were identified in our cohort: rs1134986, a missense variant in *DLG1* (NM_001204387: c.485G>A:p.R162Q,(n=1/6)), and rs939885, a variant in *SLC51A* (NM_152672: c.604G>A:p.V202I (n=2/6)), with an allele frequency of 15% and 48%, respectively, suggesting that these are most likely benign polymorphisms. In addition, the SNP data on the remaining allele revealed that the parental allele of origin of the deletion in five of six cases is the fatherâ□□s allele.

PrgmNr 3809 - Rapid Identification of Genetic Factors Contributing to Autism Spectrum Development with Disproportionate Megalencephaly in a Model System

View session detail

Author Block: **S. Nishizaki**[1], C. W. Nordahl[1], D. G. Amaral[1], M. Y. Dennis[2]; [1]Univ. of California Davis, Davis, CA, [2]Univ. of California, Davis, Davis, CA
Disclosure Block: **S. Nishizaki:** None.
Among autism spectrum development (ASD) individuals some of the worst prognoses come from comorbidity with accelerated brain growth, known as disproportionate megalencephaly (DM). However, little is known about the genetic factors contributing to ASD-DM. In this study we identified de novo variants potentially contributing to ASD-DM using trio and quad whole genome sequencing data from the MIND Instituteâ□□s Autism Phenome Project (APP) and Simonâ□□s Simplex Collection (SSC) cohorts. Through this analysis we have identified multiple genes previously associated with ASD, including *PAX5*, *SCP2*, and *ADCY5*. We are now examining the consequence of these proband loss-of-function variants on head-size phenotypes by generating knockout zebrafish using CRISPR gene-editing technology of the genes we identified as putatively associated with ASD-DM. Here, we demonstrate the utility of the VAST BioImaging System to rapidly image knockout zebrafish of a known macrocephaly-associated gene, *wdfy3*. We hope to quickly identify novel ASD-DM associated genes to further advance the identification of relevant disease pathways, potential genetic therapy targets, and early detection markers in ASD-DM.

PrgmNr 3810 - Regulating translation of the *GJB1* 5' UTR

View session detail

Author Block: **B. Grosz**[1], J. Svaren[2], G. Perez-Siles[1], G. A. Nicholson[3], M. L. Kennerson[1]; [1]ANZAC Res. Inst., Concord, Australia, [2]Univ. of Wisconsin-Madison, Madison, WI, [3]Molecular Med. Lab., Concord, Australia

Disclosure Block: **B. Grosz:** None.

Mutations of *GJB1* coding and non-coding regions cause the second most common form of Charcot-Marie-Tooth neuropathy, CMTX1. The non-coding *GJB1* c.-103C>T mutation [chrX:71,223,249 (hg38)] is in exon 1b of the *GJB1* P2 5â□□ untranslated region (5â□□ UTR) and has been reported to cause CMTX1 in multiple unrelated families. *GJB1* c.-103C>T is in an evolutionarily conserved region that is flanked by two non-pathogenic SNPs, c.-109C>T (rs746618959) and c.-102G>A (rs753207004). To investigate whether the c.-108_-103 region is a regulatory element that may be disrupted by *GJB1* c.-103C>T, a suite of luciferase constructs was generated. The *GJB1* neural-specific P2 promoter and 5â□□ UTR was inserted upstream of the firefly luciferase (FLuc) gene in the pGL4 vector, and the *GJB1* 3â□□ UTR was inserted downstream of the FLuc gene. *GJB1* c.-103C>T and *GJB1* c.-108_-103del were then introduced separately into the *GJB1*-pGL4 vector. These vectors were separately transfected into the RT4 rat Schwann cell line, using the pRLuc-TK vector as a transfection control. Relative expression was assessed by determining the ratio of FLuc:RLuc for *GJB1*-pGL4 vectors containing *GJB1* c.-103C>T and *GJB1* c.-108_-103del and comparing this to FLuc:RLuc ratios from the wild type *GJB1* vector and an empty pGL4 vector. The *GJB1* c.-108_-103del construct resulted in a 46% decrease (M=0.54, SD=0.19) in FLuc expression when compared to the wild type *GJB1* construct (p=.014). The *GJB1* c.-103C>T mutation resulted in an 88% decrease (M=0.12, SD=0.02) in FLuc expression when compared to the wild type *GJB1* construct (p<.00001 although>GJB1 c.-108_-103del resulted in a significant decrease in expression, it remained significantly different to the c.-103C>T mutation (p=.019). While this region may represent an important *GJB1* regulatory element, the loss of expression due to the *GJB1* c.-103C>T mutation should be investigated further so that a suitable therapeutic approach can be developed.

PrgmNr 3811 - Transcriptome analysis provides insights into the possible function(s) of *SPATA5L1*

View session detail

Author Block: **S. Shetty**; Univ. of Arizona, Phoenix, AZ
Disclosure Block: **S. Shetty:** None.

SPATA5L1 is a paralog of SPATA5 (spermatogenesis associated 5). Autosomal recessive mutations in SPATA5 has been associated with intellectual disability, sensorineural hearing loss (SNHL), movement disorders and epilepsy. GWAS studies have associated multiple SNPs on SPATA5L1 with chronic kidney disease but none with neurological symptoms. This study was conducted with the help of GeneMatcher and reports 28 unique SPATA5L1 variants in 47 affected individuals with neurological findings in 25 patients (18 families). Exome sequencing, neuroimaging and RNA-Seq was performed on this multi-center cohort. 25 of these variants were present in a compound heterozygous form, three were homozygous variants and all variants were rare or absent in gnomAD. Most of these patients demonstrated either spasticity (17/25, 68%) or dystonia (15/25, 60%) or a combination of both (13/25, 52%). There was significant to profound cognitive impairment in these patients. Microcephaly (13/25, 52%) and facial dysmorphism (9/25, 36%) was also observed. Neuroimaging showed reduced cortical volume, periventricular leukomalacia, widened Sylvian fissures and hypoplastic corpus callosum. 22/47 individuals with biallelic SPATA5L1 variants showed isolated sensorineural hearing loss (SNHL) and all these individuals were of Ashkenazi Jewish descent. Principal component analysis of transcriptomic data generated from fibroblast cell lines was able to discriminate SPATA5L1 affected individual from unaffected controls, with the first two principal components explaining 54% of the variance. Over-representation analysis (ORA) of significantly down-regulated genes showed clustering for Gene Ontology (GO) terms in mitosis (mitotic spindle organization, sister chromatid segregation), DNA replication (DNA conformation change, single stranded DNA binding, DNA helicase activity) and adhesion receptors, which connect cell-substrate junctions and include fibronectin-binding, integrins, cadherins, and immunoglobulin superfamily members. We thus present the first reported pathogenic variants in SPATA5L1 associated with a mixed/complex neurological phenotype, spastic-dystonic cerebral palsy, and/or SNHL from unrelated families. This study provides insights into the role of SPATA5L1 in auditory perception and brain development.

PrgmNr 3813 - Comprehensive survey of the functional impact of SNP accessible variants in *PSAT1* using a high throughput yeast assay

View session detail

Author Block: **M. Xie**[1,2], G. A. Cromie[1], R. Lo[1], K. Owens[1,2], N. J. Kutz[2], R. N. McLaughlin[1], A. M. Dudley[1]; [1]Pacific Northwest Res. Inst., Seattle, WA, [2]Univ. of Washington, Seattle, WA
Disclosure Block: **M. Xie:** None.
Loss of function mutations in genes of the serine biosynthesis pathway (*PHGDH*, *PSAT1*, and *PSPH*) cause a set of rare, autosomal recessive diseases known as Neu-Laxova syndrome (NLS) or serine deficiency disorders. The phenotypic spectrum of these inborn errors of metabolism includes severe neurological symptoms, failure to thrive, and lethality. However, timely detection of pathogenic variants can have an enormous, positive impact on patient health. Prenatal and postnatal supplementation with serine and glycine can ameliorate, and in some cases, completely prevent the onset of symptoms. To aid variant annotation for this rare and devastating disease, we developed a yeast â□□surrogate geneticsâ□□ assay in which the highly conserved human protein coding sequence for the second enzyme in the pathway, phosphoserine aminotransferase (PSAT1), functionally replaces the deletion of its yeast ortholog (*SER1*). In this assay, yeast growth in the absence of serine provides a quantitative readout of the human enzymeâ□□s function. Previous work from our group demonstrated that results from this assay agree well with clinical annotations and the disease literature. Here, we extend that analysis to present quantitative results for the functional impact of >80% (n=1986) of all single nucleotide polymorphism (SNP) accessible *PSAT1* missense variants. Our approach leverages the availability of low-cost, large-scale gene synthesis and the development of high-throughput in vivo assays of protein function (Multiplexed Assays of Variant Effect, MAVEs) in the model organism, *Saccharomyces cerevisiae*. We will discuss the extent to which these results agree with a set of recently published pathogenic variants and the spectrum of Variants of Uncertain Significance (VUS) that have been detected in the human population. We will also present our results in the context of the published PSAT1 crystal structure. Taken together, our work provides an example of the ways in which large-scale functional assays in model systems can be powerfully applied to the study of rare diseases.

PrgmNr 3814 - Development of multiplex synthetic positive controls for an expanded *CFTR* mutation testing

View session detail

Author Block: **T. E. Angeloni**[1], S. Rosenthal[1], R. E. Nakles[1], A. Gerasimova[2], D. Root[1], Y. Liu[1], H. K. Chong[1], K. Kazmierkiewicz[1], D. Xu[1], A. Bhattacharya[3], R. Owen[1], F. L. Lacbawan[4]; [1]Quest Diagnostics, San Juan Capistrano, CA, [2]San Diego, CA, [3]Quest Diagnostics, Ladera Ranch, CA, [4]Quest Diagnostics Nichols Inst, San Juan Capistrano, CA
Disclosure Block: **T.E. Angeloni:** Salary/Employment; Quest Diagnostics.
Introduction: Cystic fibrosis (CF) is one of the most common autosomal recessive conditions in Caucasians, with a prevalence of approximately 1 in 2,500 to 3,300 live births. Pathogenic variants in the *CFTR* gene cause CF. With the improvement of molecular diagnostic technologies and genetic knowledge, a *CFTR* variant panel that expands beyond the American College of Medical Genetics-recommended population-screening panel of 25 variants may be appropriate for certain diagnostic testing purposes. However, the lack of readily available positive controls beyond the core panel challenges the development, validation, and quality control of an expanded *CFTR* variant panel. In this study, we report a simple approach to generate multiplex synthetic *CFTR* indel controls, particularly valuable for development of next-generation sequencing (NGS) pipelines. **Methods**: A total of 95 *CFTR* hotspot indel mutations were curated using gnomAD, OMIM, ClinVar, HGMD, and CFTR2 database. Artificial DNA sequences containing multiplex hotspot indel mutations were created and 8 fragments of 1.3 to 2.7 kb, each with 6 to 21 indels in 2 to 8 exons with flanking intronic sequences, were synthesized. The flanking sequences were retained to serve primer binding sites and to include intronic mutations. Two control fragment pools, each with 68 and 27 indels, were prepared by diluting 3 or 5 synthetic fragments within genomic DNA background and were sequenced by validated PCR-based or capture-based NGS methods along with patient samples. **Results**: Multiplex control fragment pools generated 100% coverage for all the target regions in every trial by both PCR-based and capture-based NGS methods, equivalent performances compared to genomic DNA samples. No sign of control fragment cross contamination was observed from 8 independent set-ups for over 2,500 patient samples. All expected indel mutations were detected by both NGS methods with informatics pipelines developed in-house except c.1820_1903del84, which was only detected by capture-based NGS method. Approximately 25% (24/95) of indel mutations with complex or repetitive neighboring sequences were called at a shifted genomic coordinate, up to 5 bp upstream from the reported position by mutation databases. **Conclusions**: We described here a simple approach to generate multiplex synthetic *CFTR* indel controls that can facilitate development, validation, and quality control of an expanded *CFTR*-variant panel. The synthetic multiplex indel controls demonstrated equivalent performances compared to patient genomic DNA samples yet minimized the number of positive controls to be tested owing to the simultaneous detection of multiple mutations.

PrgmNr 3815 - Down syndrome in patients with otitis media is associated with changes in the nasopharyngeal microbiota

View session detail

Author Block: **C. Elling**[1,2], S. Haville[3], K. Tholen[3], D. Curtis[1], T. M. Wine[1,3], N. R. Friedman[1,3], K. H. Chan[1,3], D. N. Frank[4], B. Herrmann[1,3], R. P. Santos-Cortez[1,5]; [1]Dept. of Otolaryngology- Head & Neck Surgery, Sch. of Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [2]Human Med. Genetics and Genomics Program, Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [3]Dept. of Pediatric Otolaryngology, Children's Hosp. Colorado, Aurora, CO, [4]Div. of Infectious Diseases, Dept. of Med., Sch. of Med., Univ. of Colorado Anschutz Med. Campus, Aurora, CO, [5]Ctr. for Childrenâ□□s Surgery, Childrenâ□□s Hosp. Colorado, Aurora, CO

Disclosure Block: **C. Elling:** None.

Otitis media (OM) is inflammation of the middle ear (ME) that is usually due to infection. Globally, OM is the most frequently diagnosed disease in young children and infants and is a leading cause of hearing loss. Children with Down syndrome (DS) demonstrate even higher incidence rates of OM and experience more severe OM with worse outcomes, often requiring multiple surgeries. Although individuals with DS have increased susceptibility to infections, no studies to date have investigated whether bacterial commensal or pathobiont profiles of DS children with OM differ from non-DS children with OM. Greater knowledge of the microbiota changes associated with DS will aid in pinpointing which mucosal and epithelial processes that regulate the microbiotas are disrupted in DS. Using 16S rRNA gene sequencing, we examined the microbiotas of the nasopharynx (NP) and ME of 14 children with DS and10 non-DS children; all individuals had OM. Data were analyzed in terms of microbial diversity indices and relative abundance of individual taxa. In the ME, neither alpha- or beta-diversity indices were significant. DS is associated with *Finegoldia* (nominal-p=0.05), though sample sizes for both groups were very low due to the difficulties and limitations of obtaining ME samples. In the NP, DS samples had increased alpha-diversity (S_{obs} p=0.03, Shannonâ□□s Index p=0.03) with more taxa observed compared to non-DS children. Additionally, 13 taxa were associated (nominal-pActinomyces, Atopobium, Corynebacterium, Propionibacterium, Rothia, Abiotrophia, Anaerococcus, Bacillus, Lactobacillales, Streptococcus, Veillonella, Gammaproteobacteria and *Halomonas*. Of these, *Corynebacterium*, *Halomonas* and *Lactobacillales* were previously identified as commensals in the NP, whereas *Actinomyces, Propionibacterium, Anaerococcus, Bacillus, Streptococcus* and *Gammaproteobacteria* are taxa known to include OM pathogens. Additionally, in previous studies, *Rothia, Propionibacterium* and *Streptococcus* were increased in the oral and/or subgingival microbiotas of DS individuals, potentially contributing to increased incidence of gingivitis and periodontitis related to DS. Overall, in children with OM, DS is associated with increased biodiversity and higher relative abundance of specific taxa in the NP, suggesting that dysbiosis in the NP contributes to OM susceptibility in children with DS. These findings increase our knowledge of how Down syndrome influences regulation of the mucosal microbiota and contributes to OM pathology.

PrgmNr 3816 - Dynamic cell-type-specific eQTL from single-cell RNA-seq of inflamed and non-inflamed colon elucidate biological mechanisms of ulcerative colitis

View session detail

Author Block: **K. Jagadeesh**[1], K. K. Dey[2], D. T. Montoro[1], H. Shi[3], A. Nathan[4], H. Huang[5], A. Gusev[6], C. Smillie[1], S. Raychaudhuri[7], A. Regev[8], R. J. Xavier[1], A. L. Price[9]; [1]Broad Inst. of MIT and Harvard, Cambridge, CA, [2]Harvard Sch. of Publ. Hlth., BOSTON, MA, [3]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [4]Harvard Univ., Boston, MA, [5]Boston, MA, [6]Dana-Farber Cancer Inst., Boston, MA, [7]Brookline, MA, [8]Broad Inst. of MIT and Harvard, Cambridge, MA, [9]Harvard Sch Pub Hlth, Boston, MA
Disclosure Block: **K. Jagadeesh:** None.
GWAS have successfully identified thousands of disease-associated variants, but an understanding of the underlying molecular mechanisms and the cell types through which they act remains elusive. Bulk RNA-seq expression quantitative trait loci (eQTL) have provided valuable insights into gene regulation, but do not reach the fine resolution of cell types and states. Identifying cell type eQTL and dynamic cell type eQTL using scRNA-seq data has substantial potential to bridge this gap.
We analyzed scRNA-seq data from 306K cells in both inflamed and non-inflamed colon tissue from 25 genotyped individuals with or without ulcerative colitis (UC), aiming to identify cell type eQTL and dynamic cell type eQTL (specific to inflamed tissue) impacting UC risk. We observed high concordance in eQTL associations between cell type pseudobulk linear models and Poisson mixed effect models (r=0.91), and report results from pseudobulk models below. Across 9 immune, epithelial and stromal cell types, we identified 4,264 genes with significant cell type eQTL (FDRWe leveraged the cell type eQTL results to pinpoint disease genes and cell types of action at UC GWAS loci. Examples include *ANKRD55* in T lymphocytes and *PPIF* in macrophages, both of which are biologically plausible and with the eQTL SNP in strong LD with the GWAS SNP but have not previously been strongly implicated in UC disease risk. We also performed a dynamic transcriptome-wide association study (TWAS) by constructing separate gene expression prediction models for inflamed and non-inflamed tissue and contrasting their associations to UC. Genes associated with IBD specifically by the inflamed TWAS model (and further validated by MAGMA z-score > 5) include *QRICH1* (in B lymphocytes), which plays a central role in ER stress and inflammatory colon disease; *CARD9* (in B lymphocytes), which functions in gut microbe sensing; and *IL12RB2* (in fibroblasts), which contributes to the inflammatory response and host defense. eQTL fine-mapping of *IL12RB2* in inflamed fibroblasts implicated 4 strongly linked SNPs (average r^2=0.99) located 700kb upstream of *IL12RB2* (and closer to 8 other genes). Our findings highlight the potential of cell type eQTL and dynamic cell type eQTL to elucidate cell-type-specific and cell-state-specific disease biology.

PrgmNr 3817 - Functional Validation of *SOX2* Variants Contributing to Isolated GnRH Deficiency

View session detail

Author Block: **J. Cassin**[1], M. Stamou[2], K. W. Keefe[3], K. Sung[1], A. Chin[1], R. Balasubramanian[2], S. B. Seminara[2], P. L. Mellon[1]; [1]UCSD, San Diego, CA, [2]Massachusetts Gen. Hosp., Boston, MA, [3]Brigham and Women's Hosp., Boston, MA

Disclosure Block: **J. Cassin:** None.

Isolated GnRH deficiency (IGD) is a disorder characterized by low circulating sex steroids and delayed or absent puberty. Mutations in the *SOX2* gene have been previously linked to a syndromic form of IGD with additional ocular and neurodevelopmental phenotypes. The role of *SOX2* variants in non-syndromic forms of IGD remains unclear. To close this gap, we reviewed whole exome sequencing data in a large cohort of IGD (n=1453) patients ascertained by a reproductive phenotype. We identified a total of 9 heterozygous *SOX2* rare variants (3 *de novo*) contributing to both syndromic and non-syndromic forms of IGD. To determine the pathogenicity of the discovered variants, we utilized several *in* vitro methods. First, we confirmed that *SOX*2 is expressed in kisspeptin neurons *in vivo* in the mouse brain. SOX2 colocalizes with kisspeptin in the arcuate nucleus of both male and female mice and in the anteroventral periventricular nucleus in female mice. We next investigated the effect of SOX2 on kisspeptin expression *in vitro.* We showed that SOX2 binds to the human kisspeptin promoter, and represses kisspeptin luciferase expression in two immortalized hypothalamic cell lines, KTaR and KTaV. Of the nine *SOX2* variants identified, four diminished or reversed this repression. Finally, we investigated the molecular mechanism behind the phenotype of each of the four functionally deleterious mutations. Two missense mutations prevent proper localization to the nucleus. Two truncating mutations retain their ability to bind DNA and appear to act as dominant negative mutations when titrated in luciferase assays. This study demonstrates that the *SOX2*-related human disease spectrum may include IGD without severe ocular or neurodevelopmental phenotypes. We also show novel mutational mechanisms including dominant-negative effects contributing to *SOX2*-related human disease. This study greatly expands the understanding the mutational spectrum and the underlying mechanisms of the mutations leading to IGD and informs a more complete picture of the complexity of the genetic landscape governing the hypothalamic-pituitary-gonadal axis.

PrgmNr 3818 - Harnessing natural genetic variation in *Drosophila* to characterize the underlying mechanisms of ER stress preconditioning reveals role of histone methylation

View session detail

Author Block: **K. Owings**[1], C. Y. Chow[2]; [1]Univ. of Utah, Salt Lake City, UT, [2]Univ of Utah, Salt Lake City, UT
Disclosure Block: **K. Owings:** None.
Organisms face many stressors, and an ongoing challenge is understanding how an individual can respond to numerous insults over a lifetime. The accumulation of misfolded proteins results in cellular stresses, including endoplasmic reticulum (ER) stress. Many studies examine the ER stress response in isolation. In reality, cellular stresses rarely occur in isolation but often in the context of other stresses. Little is known about ER stress preconditioning, whereby conditioning with low levels of stress alters the ability to withstand subsequent ER stress. This project aims to use natural genetic variation to characterize ER stress preconditioning and its underlying mechanisms.
I began with an ER stress preconditioning screen that utilized the 200 strains of the *Drosophila* Genetic Reference Panel (DGRP). Flies were subjected to heat shock (or no heat shock control), allowed to recover, placed on tunicamycin to induce ER stress until death, and survival was measured. Different genetic backgrounds led to a striking range in phenotypic responses to ER stress preconditioning, ranging from dying half as fast to 4.5 times faster with preconditioning than with no preconditioning. A genome-wide association study revealed that histone H3-K4 methylation is a strong potential mechanism of ER stress preconditioning. Several candidate modifiers have known roles in histone methylation. H3-K4 methylation marks promoters at transcribed genes and may play a role in transcriptional regulation. These hits solidify a potential role of transcriptional regulation underlying ER stress preconditioning. RNAseq was performed in the phenotypically extreme DGRP strains at different points in the preconditioning protocol to identify potential predictive gene expression signatures. Differentially expressed genes indicate a potential role of immune genes in ER stress preconditioning.
An effective ER stress response is critical for healthy development and aging. Disruptions in this response have been implicated in multiple human diseases, from diabetes to neurodegeneration. Understanding how previous stress events influence the ER stress response will provide insight into this pathway's fundamental biology and have important implications for approaching therapeutic development.

PrgmNr 3819 - High throughput, yeast-based functional assays of genetic variants in human genes associated with inherited anemias

View session detail

Author Block: **M. Tang**, R. Lo, G. Cromie, M. Timour, J. Ashmead, A. Sirr, A. M. Dudley; Pacific Northwest Res. Inst., Seattle, WA
Disclosure Block: **M. Tang:** None.

Inherited anemias are highly actionable genetic disorders. Early diagnosis can help physicians flag drug contraindications and prescribe dietary restriction or supplementation. However, most polymorphisms detected by clinical genome or exome sequencing are Variants of Uncertain Significance (VUS), which cannot inform diagnosis. High throughput assays or Multiplexed Assays for Variant Effect (MAVEs) performed in model systems are one means of generating functional information needed for variant interpretation at scale. Functional studies in *Saccharomyces cerevisiae* can be developed and performed in less time and at lower costs than in human cell lines due to the speed and ease with which yeast can be genetically modified, robotically manipulated, and quantitatively assayed. Here, we describe the development of yeast-based functional assays for three human metabolic genes (*G6PD, PGK1 and TPI1*) that are associated with inherited anemias. The prevalence of the associated Mendelian diseases ranges from common to extremely rare. G6PD deficiency is the most common human enzymopathy, affecting an estimated 400 million people worldwide. In contrast, PGK1 deficiency has been documented in only ~40 individuals. Because these genes are highly conserved and the enzymes they encode perform the same biochemical functions, the human protein coding sequences can functionally replace (genetically complement) deletions of their corresponding yeast orthologs. As such, yeast growth in media conditions that require their activity provide quantitative readouts of the human enzymeâ□□s function. Our functional assays are robust, quantitative, and easily expandable to a scale that permits a single researcher to construct and assay all SNP-accessible amino acid substitutions across the length of the protein. Importantly, our results with a small set of well characterized disease alleles in each gene show good agreement with disease severity in the literature. Thus, our assays provide the foundation for comprehensive testing of the functional impact of amino acid substitutions in the human protein sequence of these genes. Our goal is to make data generated using these assays available and accessible to clinical geneticists performing variant interpretation.

PrgmNr 3822 - 90% of cryptic-donors activated in genetic disorders are present in variant-free RNA-Seq samples

View session detail

Author Block: **R. E. Dawes**, H. Joshi, S. J. Bryen, A. Bournazos, S. Cooper; Kids NeuroSci. Ctr., Sydney, Australia
Disclosure Block: **R.E. Dawes:** None.
Predicting which cryptic-donors may be activated by a genetic variant has proven intractable. We show rare mis-splicing events in variant-free RNA-seq samples are a highly accurate predictor of cryptic-donor selection in rare disease, through analysis of cryptic-donors activated by 4,811 variants, versus decoy-donors (any GT or GC not used). Conversely, the *a priori* utility of most splicing algorithms is limited. While *In silico* analysis accurately predicts variantsâ deleterious effect on authentic-donors (85 - 99 % weakened) or beneficial effect on cryptic-donors (67-98% strengthened), 29 - 62% of cryptic-donors do not out-compete the (variant) authentic-donor, and 34 - 67% are not the strongest decoy-donor within 250 nt. The influence of auxiliary splicing elements is likely to blame for current algorithmic insufficiency - we show G-repeats in the first 50nt of the intron can mask the presence of otherwise strong decoy-donors. Evidence of a donor's *splice-competence* from variant-free RNA-Seq samples is a potent predictor of cryptic-donor activation, with 90% of activated cryptic-donors present and 95% of unused decoy-donors absent.

PrgmNr 3823 - Automated annotation of human centromeres

View session detail

Author Block: **A. Bzikadze**[1], O. Kunyavskaya[2], T. Dvorkina[2], I. Alexandrov[2], P. Pevzner[1]; [1]Univ. of California, San Diego, La Jolla, CA, [2]Ctr. for Algorithmic Biotechnology, Inst. of Translational Biomedicine, Saint Petersburg State Univ., Saint Petersburg, Russian Federation
Disclosure Block: **A. Bzikadze:** None.

Recent advances in long-read sequencing technologies led to rapid progress in centromere assembly. Last year, we presented *centroFlye* â the first automated tool for centromere assembly from long error prone reads (Bzikadze and Pevzner, *Nature Biotechnology*, 2020). The same year, we joined efforts with the Telomere-to-Telomere Consortium that, in May 2021, announced a landmark achievement â the first *complete* assembly of the whole human genome (Miga et al., *Nature*, 2020; Logsdon et al., *Nature*, 2021; Nurk et al., *bioRxiv* 2021). This progress opened a possibility to address the long-standing questions about the architecture of human centromeres.

Centromeres are tandem repeats that are formed by units repeated thousands of times with limited nucleotide-level variations but extensive variations in copy numbers in the human population. Each unit represents a tandem repeat formed by smaller repetitive building blocks thus forming a nested tandem repeat. Partitioning all blocks into *n* clusters of similar blocks defines *n* monomers. A canonical order of monomers (referred to as a higher-order repeat or HOR) is specific for each centromere and is defined as the ancestral order of monomers that has evolved into the complex organization of extant centromeres. The current view of the centromere evolution is summarized by the *Centromere Evolution (CE) Postulate*: each centromere has evolved from a single ancestral HOR formed by different monomers. Although the CE postulate is universally accepted, we are not aware of a rigorous proof of this postulate or an algorithm that, given an extant centromere, derives its HOR. We developed *CentromereArchitect* that addresses both the monomer inference and HOR inference as two separate problems (Dvorkina et al., *Bioinformatics*, 2021). However, more recent analysis revealed that, to generate a biologically adequate centromere annotation, the monomer and HOR inference should be viewed as two interconnected problems.

We present the *HORmon* algorithm that incorporates the monomer and HOR generation into a single pipeline and generates the first automated centromere annotation that is largely consistent with the CE Postulate and previous manual centromere annotations (Uralsky et al., 2019). Recognizing that HORs represent an important evolutionary concept, we show how HORmon can be used to automatically derive the currently known HORs. At the same time, we argue that the ancestral HORs do not necessarily represent the best way to annotate the extant centromeres and introduce an alternative concept of a monorun graph. We project that it will complement the concept of a HOR in future centromere studies.

PrgmNr 3824 - Cell type deconvolution of whole blood bulk RNA-Seq to reveal biological insights

View session detail

Author Block: **T. A. Boltz**[1], T. Schwarz[1], M. Bot[1], C. Duan[1], B. Pasaniuc[2], R. A. Ophoff[3]; [1]UCLA, Los Angeles, CA, [2]Geffen Sch. of Med. at UCLA, Los Angeles, CA, [3]Univ California Los Angeles, Los Angeles, CA

Disclosure Block: **T.A. Boltz:** None.

GWAS have successfully discovered loci associated with various phenotypes, including quantitative biological measurements or risk for diseases. Given that the majority of these loci are in non-coding regions of the genome, a common approach in pursuit of discovering causal biological mechanisms is eQTL analysis. However, eQTL analysis of bulk tissue can often dilute cell-type specific signals and mask trait-relevant mechanisms, while single-cell sequencing can be prohibitively expensive in large cohorts. Here, we use bMIND to compute the cell-type specific (CTS) gene expression estimates of a bulk RNA-seq dataset from whole blood samples of 1,996 individuals, including patients with bipolar disorder and schizophrenia, as well as controls. Using the LM22 signature matrix from CIBERSORT as a reference and non-negative least squares deconvolution results in eight different immune cell types with proportions >= 0.1 within our dataset, including: neutrophils, naÃ¯ve B cells, memory B cells, CD8 T Cells, naÃ¯ve CD4 T cells, memory CD4 T cells, NK resting cells, and monocytes. These proportions were used to derive the CTS expression estimates for >19,000 genes per individual, across the eight cell types. We ran separate eQTL analyses per cell type, finding >2,500 significant (FDR 50%). We characterized these sets of eGenes per cell type using MAGMA to reveal enrichment with various traits. We find that the eGene sets from all cell types except naÃ¯ve CD4 T cells have significant (FDR

PrgmNr 3825 - Comprehensive variant calling in the *GBA* gene using whole-genome sequencing data

View session detail

Author Block: **M. A. Eberle**[1], X. Chen[1], M. Toffoli[2], F. J. Sedlazeck[3], S. Scholz[4], A. H. V. Schapira[2], C. Proukakis[2]; [1]Illumina Inc., San Diego, CA, [2]Univ. Coll. London Queen Square Inst. of Neurology, London, United Kingdom, [3]Baylor Coll. Med., Houston, TX, [4]NIH, Bethesda, MD
Disclosure Block: **M.A. Eberle:** Salary/Employment; Illumina Inc..
Mutations in the *GBA* gene cause Gaucher disease when inherited in an autosomal recessive manner. Carriers of these mutations are at an increased risk of two closely related neurodegenerative conditions, Parkinsonâ□□s disease (PD) and Lewy body dementia (LBD). The presence of a highly homologous nearby pseudogene (*GBAP1*) predisposes *GBA* to gene conversion and reciprocal recombination resulting in copy number gains or losses, as well as small pathogenic variants. Genetic testing and analysis of *GBA* are complicated due to the homology with *GBAP1* and a variety of recombinant variants that are *GBAP1*-derived, where the *GBA* sequence is replaced by the corresponding sequence of *GBAP1*.

Here, we present a novel bioinformatics method, Gauchian, that detects *GBAP1*-derived recombinant variants in the homology region as well as other *GBA* mutations based on short-read whole genome sequencing (WGS) data. Gauchian shows high performance with validation of key *GBAP1*-derived variants by Oxford Nanopore Technologies (ONT) targeted sequencing and copy number variants by digital PCR, outperforming the GATK Best Practices pipeline, as standard aligners have difficulty in accurately aligning reads derived from *GBAP1-GBA* hybrids.

Applying Gauchian to 10,623 samples from the 1000 Genomes Project and Accelerating Medicines Partnership - Parkinsonâ□□s Disease (AMP-PD) data sets reveals that copy number gains involving *GBA* and *GBAP1* are much more common and more variable in Africans, but do not appear to be associated with a risk of PD or LBD. Gauchian further confirms the elevated *GBA* mutation rate in LBD (11.8%) compared to PD (7.8%) cases. Compared with the GATK Best Practices pipeline, Gauchian doubles the number of *GBAP1*-derived variant calls in the homology region, including p.A495P, p.L483P, p.D448H and c.1263del, RecNciI and c.1263del+RecTL.

These findings highlight the importance of accurate *GBA* mutation detection across patient populations, particularly as targeted treatments are already at the clinical trials stage. Gauchian is the first short-read WGS-based tool for identifying carriers of *GBA* variants, particularly the *GBAP1*-derived variants, enabling *GBA* testing to be offered as part of a comprehensive test in carrier screening or other clinical settings where WGS is performed.

PrgmNr 3826 - Computational framework for consensus variant calling and variant call aggregation

View session detail

Author Block: **M. Samadi**[1], J. Ng[2], P. Vats[1], S. Onken[1], S. Sarkar[2], A. Sethia[1], T. Harkins[1], T. Turner[2]; [1]NVIDIA, Santa Clara, CA, [2]Washington Univ. Sch. of Med., St. Louis, MO
Disclosure Block: **M. Samadi:** None.
The development of different bioinformatics pipelines has continuously improved the variant calling accuracy for gene panels, exomes, and whole-genome sequencing; however, different variant callers can yield widely differing lists of variants from the same data set. The main reason is that each variant caller uses a different statistical method providing unique strengths and weaknesses of each approach. In order to overcome these discrepancies, recent studies show that consensus calling approaches are generally more robust and accurate than depending on an individual caller; however, running multiple variant callers and aggregating the results in a substantial increase in the required computational resources, turnaround time, along with the need to employ a complex workflow.
To overcome this challenge, we developed a computationally accelerated framework for consensus variant calling and variant call aggregation (VCA). This framework enables bioinformaticians to run multiple variant callers on the same data and gather the results in an easy and efficient way. To provide orders of magnitude faster turnaround, we optimized the whole framework to use Graphics Processing Units (GPUs) as accelerators. We have accelerated multiple variant callers on GPU such as GATK Haplotype Caller and Mutect2, Google DeepVariant, SomaticSniper, BCFtools, LoFreq, VarScan, and MuSE by orders of magnitude while maintaining output equivalence. As these tools are all optimized, users can run them all on a single GPU server in a few hours for a whole human genome. Moreover, users can add their own variant callers to the analysis, running them in conjunction with the accelerated variant callers. We also provide multiple simple VCA tools to use such as union, intersection, vote based vcf merger. Similar to variant callers, users can add their own VCAs to the framework.
As an example of this framework, we implemented a consensus caller for the Detection of de novo variants (DNVs). In this work, we used accelerated Google DeepVariant and Haplotype Caller and our customized VCA. The whole DNV workflow runs in ~1 hour providing a clear advantage over CPU-based approaches. We analyzed 602 publicly available trios from the 1000 Genomes Project as a control. We detected 445,711 DNVs, having a bimodal distribution, with peaks at 200 and 2000 DNVs. The excess DNVs are cell line artifacts that are increasing with cell passage. Reduction in DNVs at CpG sites and in percent of DNVs with a paternal parent-of-origin with an increasing number of DNVs supports this finding. These findings at this scale would not be possible without running multiple variant callers in an efficient and accelerated way.

PrgmNr 3827 - Evaluation of genome similarity for a twin family using 23andMe genotyping technology

View session detail

Author Block: **R. Huang**; Eastside Catholic Sch., Sammamish, WA
Disclosure Block: **R. Huang:** None.
Monozygotic twins, usually referred to as Identical twins, are two siblings who theoretically inherit exactly the same genetic information produced by the division of a single fertilized egg. Yet, in reality, somatic mutations may occur during cell division throughout the lifetime. Today, scientists still have a limited understanding of how much genetic variation exists between twins. This study aims at exploring this question using a case study based on the genomes of a family that includes a mother, a son, and a pair of twin daughters. The genotype information was obtained from direct-to-consumer genotyping services, namely 23andme. The results of this study provide insights into where and how mutations arise between twins. In total, 1,974 SNP genotypes, excluding â□□no calls,â□□ are compared, and only 12 genotypic differences from eight chromosomes are found. Among these chromosomes, chromosomes 11 contain the highest number of differences. In addition, SNPs found in chromosome 11 of two sets of public 23andme data downloaded from the Harvard Personal Genome Project were analyzed to compare and contrast twin and non-twin genomic differences. The result shows that non-twin data each has 83.82% and 83.64% of differences, while the twin data has only 0.36% of differences. This study not only confirms that identical twins have much higher genetic similarity than non-twins but also gives an accurate number of differences between typical twin genomes as an example for later studies. Also, by comparing motherâ□□s and children's data, we found the genotyping technology accuracy varied for each trial but the overall proportion remains within a reasonable range.

PrgmNr 3828 - Findings from the Critical Assessment of Genome Interpretation, a community experiment to evaluate phenotype prediction

View session detail

Author Block: **C. Bakolitsa**[1], S. Jain[2], G. Andreoletti[3], R. A. Hoskins[4], S. Repo[5], S. M. Fullerton[6], P. Radivojac[2], J. Moult[7,8], S. E. Brenner[1], CAGI Participants; [1]Univ. of California, Berkeley, Berkeley, CA, [2]Northeastern Univ., Boston, MA, [3]Univ. of California San Francisco, San Francisco, CA, [4]Univ. of California Berkeley, Berkeley, CA, [5]Elixir, Wellcome Genome Campus, Cambridge, United Kingdom, [6]Univ. of Washington, Seattle, WA, [7]Inst. for BioSci. and Biotechnology Res., Univ. of Maryland, Rockville, MD, [8]Dept. of Cell Biology and Molecular Genetics, Univ. of Maryland, College Park, MD
Disclosure Block: **C. Bakolitsa:** None.
Interpretation of genomic variation plays an essential role in the analysis of cancer and monogenic disease risk, and increasingly also in the context of complex disease susceptibility, with applications ranging from basic research to clinical decisions. Yet the field lacks a clear consensus on the appropriate level of confidence to place in variant â□□impactâ□□ and interpretation methods. The Critical Assessment of Genome Interpretation (CAGI, 'kÄ□-jÄ□) is a community experiment to objectively assess computational methods for predicting the phenotypic impacts of genomic variation. CAGI participants are provided genetic variants and make blind predictions of resulting phenotype. Independent assessors evaluate the predictions by comparing with experimental and clinical data. CAGI has completed five editions with the goals of establishing the state of art in genome interpretation, and of encouraging new methodological developments. Challenges have been predominantly based on human data, mirroring problems in clinical practice and much basic research. The focus has been on interpreting nonsynonymous variants, splicing variants, structural variation, whole-exomes and whole-genomes, with phenotypes ranging from molecular and cellular measurements to organismal phenotypes in inherited disease and cancer. Results from previous CAGI experiments have been described in two special issues of *Human Mutation* (https://onlinelibrary.wiley.com/toc/10981004/2017/38/9, https://onlinelibrary.wiley.com/toc/10981004/2019/40/9). Each edition of CAGI has revealed new aspects of methods, and a number of themes have emerged. Independent assessment has highlighted a strong signal in predicting the functional effects of missense variation. We have also found methods provide compelling evidence for use in clinical classification of variants. Structure-based missense methods excel in a few cases, while evolution-based methods have more consistent performance. CAGI has contributed to the interpretation of cancer variants. Interpretation of non-coding variants shows promise but is not at the level of missense. In examples using clinical data, predictors identified causal variants overlooked in the initial clinical analysis. CAGI challenges have also revealed complications in interpreting whole genomes, limitations of *in vitro* data, and a need for more sophisticated assessment strategies and experimental design. Future rounds of CAGI challenges will seek to incorporate larger datasets, explore polygenic models, and continue to assess the state of the art. Detailed information about CAGI may be found at https://genomeinterpretation.org.

PrgmNr 3829 - Full lifecycle continuous workflow-enabled variant benchmarking

View session detail

Author Block: **J. Leipzig**, J. Schnebly, J. Rudolph, C. Brown, K. Sebby; Truwl, Whitefish, MT
Disclosure Block: **J. Leipzig:** None.
Performance testing and benchmarking are critical prerequisites to the development of sequencing-based tests. The substantial demand for these capabilities is underserved and many groups struggle to re-implement these functionalities in-house. Particularly of high interest is pipelines for variant calling. These tools have received renewed interest with the advent of machine learning and deep learning approaches that promise to push the limits of sensitivity and accuracy in genomic analysis. Truwl enables performance testing with easy-to-use dashboards of metrics and benchmarks that allow any user to evaluate and iterate on their analysis pipelines. Truwl hosts various bioinformatics pipelines whose inputs can be defined through the platformâ□□s web-based input editor and then executed directly on the cloud. Multiple tool and parameter choices are available for sequence alignment and variant calling. For each executed job the system tracks job run statistics, parameter settings, and all inputs and outputs. Results from variant calling pipelines can be fed directly into a previously existing small-variant benchmarking pipeline that we refactored into the Workflow Description Language. This pipeline uses hap.py and bcftools to generate metrics from comparisons of VCF files to known truth-sets over user-specified regions. We have compiled reports and visualizations generated by the pipeline into notebooks that display directly in the web browser. Unique to this system is an interactive dashboard specifically designed to compare runs across truth sets, tool choices, and parameter distributions. The interface and internal implementation borrow concepts of model and hyperparameter exploration tools from the machine learning community, namely MLFlow. A high-level dashboard quickly allows users to display subset columns among the dozens of manipulated parameters and relevant metrics and between facets of the benchmarking analysis such as whole-exome vs coding region. This dashboard enables a data-driven approach to pipeline selection, optimization, and validation and can also be used in production environments to validate that pipelines and assays continue to perform properly.

PrgmNr 3830 - Length Variability in Short Tandem Repeats and Its Applications

View session detail

Author Block: **P. Vijay**[1], S. Batalov[2], Y. Ding[1], S. F. Kingsmore[3], S. Chowdhury[4], C. A. Hobbs[5,6], M. N. Bainbridge[7]; [1]Rady Children's Inst. for Genomic Med., San Diego, CA, [2]Rady Childrens Inst., Genomic Med., San Diego, CA, [3]Rady Children's Inst. for Genomic Med., Coronado, CA, [4]Rady Children's Inst., Genomic Med., San Diego, CA, [5]Univ. of Arkansas for Med. Sci., Little Rock, AR, [6]Rady Children's Inst., Genomic Med., San Diego, CA, United States, San Diego, CA, [7]Rady Children's Hosp., San Diego, CA
Disclosure Block: **P. Vijay:** None.
Short Tandem Repeats have been found to play a role in a myriad of complex traits and genetic diseases. Using GangSTR, an algorithm for genome-wide genotyping of short tandem repeats, we examined the variability in the lengths of over 800,000 STR loci under four different scenarios. First, we analyzed the variability in the length of STRs across six separate genetically determined ethnic groups: Africans, Europeans, East Asians, Mixed Americans, homogenized Americans, and Pacific Islanders. Europeans were compared with the other five ethnicities and STRs with significantly (Absolute Z-Score > 5) different lengths were identified. Mixed Americans and Pacific Islanders presented with the fewest number of STRs with significantly altered lengths, with 623 and 1560 positions respectively, while East Asians, at 3606 positions, and Africans, at 4933 positions, showed the highest number of STR loci with differing lengths. Secondly, we focused on repeat expansions at known pathogenic loci. *TCF4, AR*, and *DMPK* were the STR loci with the highest number of patients with expansions. Forty-nine and 6 pathogenic expansions were detected in TCF*4* and *DMPK*, respectively. Although 162 individuals presented with expansions in the *AR* gene, none of these patients contained expansions greater than the pathogenic length for this locus. Next, we focused on individuals with divergent STR lengths in STRs that appear highly stable across all ethnic groups. We observed long, divergent STR expansions in both known disease and non-disease genes and in coding and non-coding regions of these genes. STR expansions were observed in the protein coding regions of two genes, *ZBTB4* and *SLC9A7*, the latter of which is the cause of Intellectual developmental disorder, X-linked 108. Finally, we identified *de novo* expansions occurring in our proband cohort. One hundred eighty-four parent-child trios were identified and allele lengths were compared between proband and both parents. Alleles where the probandâ□□s length was greater than two times that of either parent were flagged as a *de novo* expansion, while probands with allele lengths shorter than the shortest allele length of either parent were flagged as a *de novo* contraction. In total, we observed 3179 potentials *de novo* expansions and just 9 *de novo* contractions with 297 expansions associated with disease causing genes and 0 contractions associated disease causing genes.

PrgmNr 3831 - Maverick: Variant prioritization for Mendelian diseases using deep learning

View session detail

Author Block: **M. C. Danzi**[1], S. Zuchner[2], M. Dohrn[1], S. Fazal[1], A. P. Rebelo[2], B. Aaron[1], V. Cintra[1]; [1]Univ. of Miami, Miami, FL, [2]Univ Miami, Miami, FL
Disclosure Block: **M.C. Danzi:** None.
Thousands of genes have been discovered to cause Mendelian-inherited diseases, according to current data from OMIM. Yet, a sizable proportion of patients with many Mendelian disorders do not currently receive a genetic diagnosis. Identifying the causal variants in these unsolved cases is an important and challenging task. We developed MAVERICK, a transformer-based neural network, to differentiate among Mendelian disease-causing dominant variants, recessive variants and all others. MAVERICK is able to classify non-synonymous SNVs and protein altering indels simply by comparing the referent and altered amino acid sequences. Patient exomes typically contain hundreds of rare, protein altering SNVs and Indels. We developed MAVERICK to filter those results down to a more manageable number of probable causal variants. Using only genotype information to rank all the variants in a patient, this approach gave the causal variant top prioritization over 77% of the time for known disease genes and over 50% of the time for novel disease genes. Furthermore, the causal variant fell within the top five ranked variants over 90% of the time for both known and novel disease genes. In these test cases, the method considered both dominant and recessive inheritance patterns and scored all possible monogenic compound heterozygous pairs in addition to single heterozygous or homozygous variant solutions. MAVERICK additionally supports the incorporation of phenotype information as HPO terms and inheritance information, which can often improve performance considerably. Here, we introduce MAVERICK: a Mendelian approach to variant effect prediction. We believe this approach will greatly reduce the time and effort required to solve cases with novel variations and even help elucidate new disease genes. MAVERICK is currently in open beta testing on the GENESIS platform, which hosts genomic data from over 11,000 Mendelian disease patients.

PrgmNr 3832 - Predictive Interpretation and Scoring Model (PrISM) increases efficiency of variant classification by reducing manual review time

View session detail

Author Block: **T. Smart**, N. Sanapareddy, E. Choo, B. Gall, I. Rakova, D. Keen-Kim; Natera, Inc., San Carlos, CA
Disclosure Block: **T. Smart:** Major Stockholder/Ownership Interest; Natera, Inc..
Salary/Employment; Natera, Inc..
Variant classification is a stepwise, resource-intensive process that incorporates the ACMG guidelines to accurately curate and classify genetic variants. Doing so requires variant curator expertise to provide a thorough review of literature, clinical and experimental evidence, databases, and other evidence sources, which is a manual and time-consuming process. In addition, periodic reassessment of pathogenic, likely pathogenic, and variants of uncertain significance (VUSs) requires time, further increasing manual review time. To improve the efficiency of our current variant classification workflow, we propose the Predictive Interpretation and Scoring Model (PrISM) for variant classification. Whereas other proposed models focus on an automated modelling process, which can skip expert review altogether, PrISM improves variant classification by gathering evidence obtained through an automated workflow to assign a confidence score to variants that have previously undergone manual review. Variants with higher confidence scores are assigned a final classification, whereas variants with a lower confidence score require manual review before a final classification is assigned.At the core of our PrISM model is a naive decision tree, incorporating several evidence categories from the manual review workflow, which was developed to identify high confidence, autosomal recessive variants. If positive (likely pathogenic/pathogenic), variants can qualify for single review, and if negative (VUSs), variants can be exempted from manual review. Evidence categories used to determine confidence scores include previous classification, concordance with a consensus ClinVar assessment, and new literary or clinical evidence. To determine the potential reduction in manual reviews achievable by PrISM, we performed a retrospective analysis calculating the number of variants that would qualify for reduced manual review. Of the total 159,553 unique variants that we previously assessed in our HorizonTM carrier screening database, 2,768 pathogenic, 547 likely pathogenic, and 9,003 VUSs qualified for either single review or exemption from manual review. The analysis included all pathogenic, like pathogenic, and VUSs existing in the database before May 2021.In this dataset, implementing our model would potentially reduce manual review for 12,318 variants, which is 7.7% of all unique variants, allowing for a faster laboratory turnaround time. Therefore, implementation of PrISM provides a low-risk method of increasing variant classification efficiency of previously reviewed variants without compromising the stringency of the manual review process.

PrgmNr 3833 - Small indel detection within VNTRs using short-read sequencing data

View session detail

Author Block: **J. Park**, M. Bakhtiari, V. Bafna; Univ. of California San Diego, La Jolla, CA
Disclosure Block: **J. Park:** None.

Motivation: Tandem repeats (TRs) are highly polymorphic regions in DNA where approximated motifs are replicated in tandem. In the human genome, there are more than one million TRs accounting for significant genetic variation. The most common variation at TRs is repeat copy number variation, and over 50 diseases are known to be caused by repeat expansion. Another type of variant is small indels within repeat units, and a couple of diseases are caused by indels in variable number tandem repeats (VNTRs) such as mendelian disorder medullary cystic kidney disease type 1 (MCKD1) and monogenic type 1 diabetes. However, small indels are relatively unexplored mainly due to the long and complex structure of VNTRs with multiple repeat units.
Methods: We extended a genotyping tool for VNTRs, adVNTR, to handle the long and complex VNTRs with short-reads. Unlike the previous version, adVNTR allows multiple repeat units to be individually modeled in VNTR allowing for more accurate capture of repeat variations. A speed-up technique was used to improve the genotyping process.
Results: In a simulated dataset, adVNTR outperformed GATK with 100% median accuracy for 20 simulation sets. We also genotyped 2237 VNTRs in coding regions using 1000 Genomes dataset and found 265 common VNTRs (occurring in > 1% of the population with non-reference allele length). Out of 265 common VNTRs, 232 (87.5%) have repeat unit length with multiple of three implicating in-frame variants. Finally, we tested our tool on 3 MCKD1-positive samples caused by a frame-shift insertion in MUC1 VNTR and successfully identified the known pathogenic variants demonstrating the ability of adVNTR to detect small indels in long and complex VNTRs.

PrgmNr 3834 - Systematic evaluation and Improvements for trans-eQTL Detection Methods Allows Identification of Novel trans-eQTLs in the GTEx data

View session detail

Author Block: **C. Wu**, T. Xu, A. Goren, M. Gymrek; Univ. of California, San Diego, San Diego, CA
Disclosure Block: **C. Wu:** None.
Studies of expression quantitative trait loci (eQTLs) have aimed to discover genetic variants that explain variation in gene expression levels due to associations with complex traits and human diseases. While thousands of cis-eQTLs have been reliably identified, consistently replicating trans-eQTL effects proved to be challenging due to insufficient statistical power, lack of comparable tissues and cohorts, and putative false positive associations. In particular, technical covariates lead to a substantial variation in expression datasets and result in multiple false positive eQTL calls. As a result, biological mechanisms and characteristics of trans-eQTLs remain largely unknown.
Here we present x-qtl â a novel trans-eQTL detection method. For a given trans-eQTL, we assume the distribution of association statistics to be a mixture of two distributions â for the target and non-target genes. x-qtl addresses two critical limitations of existing methods. First, the method allows downstream characterization of trans-eQTLs by predicting the number of target genes. Second, x-qtl applies principal component analysis to the expression matrix to account for gene correlations. Specifically, we represent the variables (genes) as the uncorrelated principal components which reduces noise from correlated genes that can obfuscate true trans-eQTL signals.
Next, we develop a simulation framework to evaluate trans-eQTL detection tools. Our framework enables for varied effect size distribution and number of target genes for each trans-eQTL, includes pairwise correlation between expression of different genes, and simulates effects of technical covariates. We use our framework to evaluate the power of x-qtl and two existing eQTL frameworks, MatrixeQTL and CPMA, to detect trans-eQTLs with a variety of characteristics. Our results indicate that only trans-eQTL with extremely large effect sizes, or affecting hundreds to thousands of target genes, can be reliably detected. Remarkably, x-qtl is more sensitive in detecting trans-eQTLs with a small number of targets than existing methods.
We benchmark x-qtl against existing tools using yeast expression data containing documented eQTLs. Then, we apply all methods to the recently released Genotype-Tissue Expression (GTEx) project v8 dataset with 838 human donors and 15,201 samples from 52 tissues. We observe a putative trans-eQTL in Nerve-Tibial involved in signal transduction pathways and another in Thyroid that is located in a thyroid-specific transcription factor. Additionally, we detect novel trans-eQTLs that are replicated among different tissues but were missed by traditional trans-eQTL analysis methods.

PrgmNr 3835 - Towards evidence-based recommendations and protocols for the use of computational tools in missense variant pathogenicity classification

View session detail

Author Block: **V. Pejaver**[1], M. S. Greenblatt[2], R. Karchin[3], K. A. Pagel[3], S. D. Mooney[1], A. O'Donnell-Luria[4,5], S. Tavtigian[6,7], B-J. Feng[6,7], L. G. Biesecker[8], S. Harrison[4], A. Byrne[4], P. Radivojac[9], S. E. Brenner[10], ClinGen Sequence Variant Interpretation Working Group; [1]Univ. of Washington, Seattle, WA, [2]Univ. of Vermont, Burlington, VT, [3]Johns Hopkins Univ., Baltimore, MD, [4]Broad Inst. of MIT and Harvard, Cambridge, MA, [5]Boston Children's Hosp., Boston, MA, [6]Univ. of Utah, Salt Lake City, UT, [7]Huntsman Cancer Inst., Salt Lake City, UT, [8]Natl. Human Genome Res. Inst., Bethesda, MD, [9]Northeastern Univ., Boston, MA, [10]Univ. of California, Berkeley, Berkeley, CA
Disclosure Block: **V. Pejaver:** Grant/Contracted Research Support (External); National Institutes of Health. Salary/Employment; University of Washington.
Current ACMG/AMP guidelines for interpreting sequence variants permit the use of computational (in silico) predictors as supporting evidence in assessing pathogenicity. However, these guidelines lack quantitative support and leave clinicians and scientists without standardized rules in applying them (e.g., PP3 and BP4 guidelines ask for consensus of two uncalibrated predictors, an approach not justified by evidence). We believe the recommendations for PP3 and BP4 should be replaced by an evidence-based, quantitative strategy. We investigated quantitative revisions of these rules.
We evaluated fourteen tools (BayesDel, CADD, EA, hEAt, FATHMM, GERP++, MPC, MutPred2, PhyloP, PolyPhen-2, PrimateAI, REVEL, SIFT, VEST4), selected based on mention in the 2015 ACMG guidelines, availability of training data, performance in the Critical Assessment of Genome Interpretation, community adoption, and balance between primary models and meta-models. We assembled a data set that excluded variants in any of these toolsâ training sets, comprised of 5,762 pathogenic or likely pathogenic (P, LP) variants and 6,208 benign (B) or likely benign (LB) variants from ClinVar. Performance criteria and score thresholds for each tool were derived through a novel approach that first calculates local posterior probabilities (Post_P) for each score and then defines score intervals based on Odds of Pathogenicity (OddsPath) criteria. We validated thresholds for each tool using an independent data set from ClinVar (3,078 P/LP and 13,531 B/LB variants).
Modeling of ACMG rules provides specific thresholds of OddsPath for Supporting, Moderate, and Strong evidence. A consensus approach using developer-defined thresholds for SIFT and PolyPhen-2 reached Moderate level of evidence for PP3 and Supporting level for BP4 for an average variant with a consensus prediction of P/LP. Because several issues raised questions about clinical application, we then selected new thresholds to account for confidence intervals around point estimates of the Post_P of pathogenicity. These thresholds apply to all (not just average) variants achieving a given level of support for pathogenicity or benignity and are more stringent than developer-defined ones. We found that all tools achieved Supporting level evidence for both pathogenic and benign classification using these data-driven thresholds. Importantly, at appropriate score thresholds, in silico methods can also provide Moderate and Strong evidence levels. We conclude that our approach and these quantitative criteria can improve confidence for in silico prediction tools as evidence for pathogenicity in variant interpretation.

PrgmNr 3836 - Variant Calling of 50,000 UK Biobank Exomes in a Weekend

View session detail

Author Block: **A. Sethia**[1], O. Krasheninina[2], T. Polanco[2], T. Zhu[1], S. Onken[1], M. Samadi[1], W. Salerno[2], T. Harkins[1]; [1]Nvidia, Santa Clara, CA, [2]Regeneron Genetics Ctr., Tarrytown, NY
Disclosure Block: **A. Sethia:** Salary/Employment; Nvidia.
The promise of human genetics and precision medicine has exploded the number of DNA samples that are analyzed every year. It is common to find studies with cohort sizes of tens to hundreds of thousands of samples. At these scales, even minor inaccuracies can occur in large enough numbers to impact downstream analyses. To maximize insight from such datasets, deep learning methods such as Googleâ□□s DeepVariant are being applied to these large datasets leading to higher accuracy of variant calling. Analysis of such a large number of samples, combined with the computational complexity of deep learning based variant callers create significant computational burden.
To overcome this challenge, we accelerate DeepVariant to run on GPUs. GPUs are renowned for their complete computing stack from hardware to software, to accelerate computation time, while increasing throughput for deep learning applications. We integrate these hardware and software technologies in the GPU accelerated version of DeepVariant to accelerate deep learning inference using a custom model. Furthermore, we accelerate the data processing workflow in DeepVariant that does not use deep learning to run in an accelerated fashion using GPUs.
We evaluated the accelerated DeepVariant using 100 samples from the UK Biobank by running the out-of-the box DeepVariant compared to the GPU accelerated DeepVariant. While the standard DeepVariant takes 58 minutes to process an exome, the GPU accelerated DeepVariant can process the same dataset in less than 5 minutes. This ~12 fold acceleration results in over 2.5 times lower cost of analysis. Across the 100 samples, 7.5 million sample level variants were called with 100% concordance between the two instances of DeepVariant. We found 1 zygosity mismatch and MEAN GQ difference of 0.43 on Phred scale between the two methods across all 100 samples.
The custom DeepVariant model shows 25% and 20% reduction in mendelian SNP and INDEL error rate respectively for 281 trios in the UK Biobank data compared to the native DeepVariant WES model.
Using the GPU accelerated variant caller using the custom model, we could analyze 50,000 UK Biobank exomes in 42 hours by using 100 spot GPU instances on Amazon Web Services (AWS). By using spot instances, we could process all the 50,000 samples for $4,800. Due to the small time it needs to run the variant calling, only 1.5% of spot instances failed and were restarted.

PrgmNr 3837 - VizCNV: A tool for orthogonal CNV validation using whole genome sequence data

View session detail

Author Block: **M. Mehaffey**[1], H. Du[2], M. Ghandi[1], C. M. Grochowski[3], K. Park[1], J. Eisfeldt[4], M. Pettersson[5], B. P. Suter[6], S. N. Jhangiani[3], A. English[7], D. M. Muzny[3], J. E. Posey[7], A. Lindstrand[5], H. Y. Zoghbi[8], J. R. Lupski[9], D. Pehlivan[7], F. J. Sedlazeck[3], Z. H. Coban Akdemir[10], C. M. Carvalho[1,1]; [1]Pacific Northwest Res. Inst., Seattle, WA, [2]BCM, Houston, TX, [3]Baylor Coll. Med., Houston, TX, [4]Karolinska Inst, Stockholm, Stockholm, Sweden, [5]Karolinska Inst.t, Stockholm, Sweden, [6]Quintara BioSci.s, Albany, CA, [7]Baylor Coll. of Med., Houston, TX, [8]Baylor Coll. of Med./HHMI/NRI, Houston, TX, [9]Baylor Col Med, Houston, TX, [10]UTHlth.Sch. of Publ. Hlth., Houston, TX

Disclosure Block: **M. Mehaffey:** None.

Copy number variation (CNV) is a significant contributor to human disease traits and genomic disorders. To date, most CNV data have been obtained using comparative genomic hybridization arrays (aCGH) or SNP arrays and as a result, most tools are designed to be compatible with those data. With the emergence of whole genome sequencing (WGS) in clinical diagnosis, identifying and characterizing pathogenic or risk-alleles for thousands of CNV calls from WGS is a significant challenge. Here, we present a visualization tool that facilitates identifying CNV events from WGS data. This tool, VizCNV, inputs WGS data and outputs paired plots of variant frequency from the genome vcf (gvcf) file and copy number variation in aCGH format based on read depth from the BAM file. Reads with low mapping quality are masked to reduce false positive calls. To test the utility of VizCNV, we analyzed data from 19 WGS trios (average coverage > 30x), 16 from an aCGH positive for *MECP2* duplication syndrome cohort and three families with probands carrying marker chromosomes. These consisted of nine *de novo* and 10 inherited CNVs, one of which contained a triplication and another with a mosaic CNV. B-allele (alternate allele) frequencies from gvcf files were extracted and the inheritance pattern inferred which established the chromosome of origin in *de novo* CNVs. Read depth was extracted from the corresponding BAM files and $\log_2$ ratios were calculated to create matching aCGH plots. The resulting read depth plots accurately displayed the CNVs identified in aCGH plots for all 19 probands. In contrast, combining output from three structural variant/CNV callers, Manta, Delly, and Lumpy, just 9/19 of aCGH observed events were called. In addition, all inheritance patterns were evident in the read depth plots, the triplication event was detected by expected $\log_2$ ratio, and the mosaic proband registered at 27% in VizCNV read depth plots which corroborates the mosaic findings in the array and FISH results. Serendipitously, because the WGS gvcf gathered information at a nucleotide level, zoomed in plots were useful to map the breakpoint junctions. The corresponding B-allele plots provided confirmation of the inheritance mode for a CNV event, if it had occurred *de novo*, or if it was mosaic in 100% of the trios. The current method of CNV calling from WGS data is still relatively inaccurate with low sensitivity and specificity. This tool is a complementary way to analyze WGS data and has close to 100% correlation with actual aCGH plots for the same sample. We propose to use VizCNV for WGS data analysis as an orthogonal validation of CNV calls and potentially a replacement for arrays for diagnostic purposes.

PrgmNr 3838 - Benefits of Hybrid Capture for Viral Surveillance of *SARS-CoV-2*

View session detail

Author Block: **D. Antaki**, T. Liu, K. Butcher, B. HÃ¶glund, J. Challacombe, E. Toro; Twist BioSci., San Francisco, CA
Disclosure Block: **D. Antaki:** Salary/Employment; Twist Bioscience.
Since its immediate outbreak in late 2019, the SARS-CoV-2 (SCV-2) virus has led to an abrupt interruption to the way of life for communities all over the world. As the virus circulated from one continent to another, novel mutations created more virulent and deadlier strains. To combat these novel strains with therapies and vaccines, researchers have begun relying on viral genome sequencing. A widely used method for targeted viral genome sequencing is amplicon sequencing, which relies on PCR amplification of short fragments. Although cost-effective, amplicon sequencing is not robust to mutations when they occur in primer sequences. However, hybrid capture, another method used for targeting the viral genome, can tolerate mismatches better resulting in fewer dropouts as viruses accumulate mutations. In this work, we provide an in silico model to predict dropouts based on mutations found in the ARTIC amplicon sequencing primer sets and the probes found in the SARS-CoV-2 NGS Assay.

Throughout this process, we aligned 383,656 high-quality genome sequences belonging to variant of concern (VOC) or variant of interest (VOI) strains in GISAID. We then profiled mismatches in ARTIC V3 primers and SARS-CoV-2 NGS Assay probes and identified isolates that would lead to diminished sequencing coverage. We detected 101,432 viruses (27%) with >=1 mismatch in the last 6 base pairs of the 3â end of ARTIC primers; of these, 413 had >=2 mismatches in one primer. In contrast, only 38 viruses (0.01%) had enough mutations (>=10) in a hybrid capture probe to have a similar effect on coverage.

Using these results, we synthesized 11 viral sequences with mutations in ARTIC primers and Twist probes. We observed that mutations in ARTIC primers led to complete dropout of the amplicon for 4 isolates and diminished coverage in an additional 4. Twist probes showed uniform coverage with little to no dropouts. Using control material (Wuhan-1 isolate Twist Bioscience) the SARS-CoV-2 NGS Assay covered 99% of the SCV-2 genome at high titers and 93% at low titers in contrast to 92% and 38% for high and low titers respectively for ARTIC amplicon sequencing.

We have shown that amplicon sequencing is limited by mutations, posing a serious concern for viral surveillance. Therefore, hybrid capture poses a better solution for variant calling and monitoring of viral strains and is more resilient to mutations as they arise. Our results establish the research only use workflow of the SARS-CoV-2 NGS Assay as a comprehensive method for viral surveillance and demonstrates the benefits of using hybrid capture vs. amplicon sequencing in order to continue to monitor for mutations found in the SCV-2 viral genome, outbreaks, and novel VOC.

PrgmNr 3839 - First Report of Spatial Whole Transcriptome Profiling of Histological Structures of Multiple Organs

View session detail

Author Block: **J. M. Beechem**[1], Y. Liang[1], A. Vanshoiack[1], W. Yang[1], L. Zhang[1], D. Patrick[1], J. Reeves[1], E. Piazza[1], M. Grout[1], C. Glaser[1], D. Kroeppler[1], E. Brown[1], K. Ton[1], M. Hoang[1], G. Demirkan[1], S. Ranade[1], K. Fuhrman[1], M. D. Rhodes[2], S. Warren[1]; [1]NanoString Technologies, Seattle, WA, [2]Nanostring Technologies, Seattle, WA

Disclosure Block: **J.M. Beechem:** Major Stockholder/Ownership Interest; NanoString Technologies. Salary/Employment; NanoString Technologies.

For thousands of years, our understanding of an organâ□□s architecture and function was based on careful visual characterization of the various tissues and substructures. Over the past decades, advances in molecular profiling techniques have increased our understanding of the underlying genetic, biochemical, and cellular factors that govern tissue organization and function. However, those molecular techniques largely relied upon dissociating or digesting the tissue, which irreparably disrupts the connection of the molecular target of interest and its physical location within the tissue. Furthermore, when the target-location information was preserved, it could only be done simultaneously for a handful of targets. In the last few years, the advent of spatial biology has finally broken that barrier, and new tools and methods are available for high plex profiling of targets in situ. For the first time, it is now possible to generate a complete understanding of the organized transcriptional patterns that define cells and tissues comprising the organs and confer their function. To realize the potential of spatial profiling to define normal organs at the cellular and molecular level, we have started an ambitious project to map the architecture of normal tissues using spatially resolved whole transcriptome profiling. In this pilot phase of the project, 5 organs (kidney, pancreas, brain, colon, and lymph node) were characterized with the GeoMxÂ® Digital Spatial Profiler using region of interests (ROI) in the key multicellular structures that confer organ function (e.g. pancreatic islets). ROI consists of geometric areas centered on the functional unit or cell type-enriched ROIs to characterize particular cell populations within the functional units. Spatial transcriptomic data are then benchmarked to orthogonal platforms as independent validation of the results. Multiple ROI per tissue and multiple donors per organ have been profiled to permit evaluation of the diversity within and between individuals. All data will be made publicly available as a resource for the community. Future expansion of this program will encompass additional organs and larger cohorts for each organ to increase the utility of this resource.

PrgmNr 3840 - Optical Genome Mapping for High Throughput Analysis of Repeat Expansion Disorders

View session detail

Author Block: **D. Zhang**[1], E. Lam[1], J. Wang[1], T. Wang[1], J. Lee[2], B. Ramandi[1], N. Miller[1], A. Pang[1], H. Sadowski[1], A. Hastie[2], M. Oldakowski[1]; [1]Bionano Genomics, San Diego, CA, [2]BioNano Genomics, San Diego, CA

Disclosure Block: **D. Zhang:** Salary/Employment; Bionano Genomics.

Expansions and contractions of simple sequence repeats are associated with more than 40 diseases including Huntingtonâ□□s disease, myotonic dystrophy, facioscapulohumeral muscular dystrophy (FSHD), Friedreichâ□□s ataxia and Fragile X Syndrome, the most common heritable cause of intellectual disability. Expanded trinucleotide repeats are the most frequent however more recently tetra-, penta-, hexa- and dodeca-nucleotide expansion have been identified as being causal in human disease. Expanded repeats are unstable and may expand during intergenerational transfer and associated disorders tend to increase in severity with each successive generation. Most repeat disorders have no direct treatments. Phenotype severity is often correlated with the amount of pathogenic expansion or contraction. Thus, accurate sizing of the repeats is crucial. Southern blotting is the gold standard for analyzing pathogenic repeats. The repetitive and polymorphic nature of these regions presents difficulties for both polymerase chain reaction (PCR), where the polymerase is unable to traverse through long repeats, and sequencing based methods which face limitations in read lengths Optical genome mapping (OGM) with the Bionano Genomics Saphyr platform has the potential to address these shortcomings. The Saphyr platform images DNA molecules up to a megabase in size labeled at specific sequence motifs in nanochannel arrays. Molecules > 150kbp are analyzed through de novo assembly or aligned to a reference genome for structural variant detection. OGM can detect germline SVs >500 bp in size including repeat expansions/contractions, inversions, reciprocal translocations, and regions with absence of heterozygosity (AOH) (>10 Mbp). Bionano Genomics has developed targeted analysis workflows for FSHD and Fragile X Syndrome. To evaluate the capability of detecting disease causing repeat expansions, we analyzed the *FMR1* repeats relevant to Fragile X syndrome using Coriell cell lines with known repeat sizes and control samples. We observed the expected expansion alleles in the Coriell cell lines, with sizes consistent with annotation and with the largest expansion being almost 1000 copies. The control samples had repeats below the pathogenic cutoff. The single-molecule data also allowed us to analyze the repeat allele spectrum, which can be indicative of the stability of the alleles. We also analyzed the DZ4Z repeat on chromosome 4 for FSHD. In addition to the canonical repeat contractions, OGM detected structural variants and other rearrangements in the vicinity of the D4Z4 locus.

PrgmNr 3841 - Simultaneous isolation of high-quality RNA and DNA from post-mortem human central nervous system tissues for omics studies

View session detail

Author Block: **N. Grima**, L. Henden, O. Watson, I. P. Blair, K. L. Williams; Ctr. for Motor Neuron Disease Res., Faculty of Med., Hlth.& Human Sci., Macquarie Univ., Sydney, Australia
Disclosure Block: **N. Grima:** None.
Post-mortem human central nervous system (CNS) tissue has garnered growing attention from multi-omics approaches, particularly for the investigation of the pathological processes underlying neurodegeneration. While most genes linked to neurodegenerative disease are ubiquitously expressed, pathology tends to occur in discrete CNS regions, making region-specific analyses of interest. Paramount to the accurate and reliable analysis of transcriptomic and epigenomic data sets in parallel, is the co-isolation of high-quality DNA and RNA. This is particularly pertinent in the case of post-mortem CNS tissues where sample availability is limited and where regional heterogeneity may mask correlations between different omics datasets. High lipid content is recognised to complicate RNA isolation from CNS tissues and consequently, existing literature has exclusively used organic solvent-based reagents and kits due to their ability to separate nucleotides and lipids into distinct fractions. In contrast, commercially available DNA/RNA co-purification kits employ guanidinium thiocyanate-based lysis in combination with column purification, forgoing use of hazardous organic solvents. It is yet to be reported whether co-purification kits can isolate high-quality nucleotides from post-mortem CNS tissues. We have developed a detailed strategy for the simultaneous isolation of high-quality DNA and RNA from post-mortem human CNS tissue. Tissue from motor cortex, frontal cortex, hippocampus, occipital cortex, anterior cingulate cortex, cerebellum and spinal cord were obtained from 22 individuals diagnosed with motor neuron disease and 13 neurologically normal controls (n = 245 tissues). We demonstrated that a DNA/RNA co-purification kit consistently isolated DNA and RNA of high yield and quality from all six brain regions. Importantly, we determined that an organic solvent-based methodology was essential to isolate RNA from spinal cord tissue. Quality assessment and RNA sequencing confirmed that the isolated RNA was suitable for downstream omics analysis. Furthermore, cerebellum RIN and tissue pH were demonstrated to be predictors of RNA integrity across all CNS regions, making them useful quality markers for future cohort selection.

PrgmNr 3842 - Single cell RNA sequencing and binary hierarchical clustering identify distinct normoxia and hypoxia associated interstitial macrophage cell types in mice exposed to hypoxia

View session detail

Author Block: **N. Campbell**[1], C. Michael[1], I. Campbell[2], S. Kumar[1], T. Phang[1], R. Kumar[1], S. Gu[1], S. Riddle[1], V. Kheyfets[1], D. Brown[1], B. Graham[3], K. Stenmark[1]; [1]Univ. of Colorado, Anschutz Med. Campus, Aurora, CO, [2]Stanford Univ., Stanford, CA, [3]Univ. of California, San Francisco, San Francisco, CA
Disclosure Block: **N. Campbell:** None.
Lung interstitial macrophages (IMs) are acknowledged to contribute to pulmonary hypertension (PH), yet few studies have examined their molecular phenotypes at the single-cell level. Therefore, we aimed to determine IM diversity and its association with PH, rationalizing that single-cell RNA sequencing (scRNAseq) and binary hierarchical clustering (BHC) within and across IMs would provide the highest single-cell resolution. Moreover, we hypothesized that these methods could identify IM heterogeneity and its association with PH. $Cx3cr1^{GFP+}$ reporter mice were exposed to normoxia (~21% FiO_2), one day (D1), or seven days (D7) of hypoxia (~10% FiO_2). $Cx3cr1^{+}$ IMs were isolated by flow cytometry. We used the 10X Genomics platform for scRNAseq, and the Cell Ranger software (v1.2.0) for alignment to the mouse genome, quality control, and filtering, resulting in n = 5144 IMs. Seurat3, ClusterMap, and Monocle2 were used for the downstream analyses. Ingenuity Pathway Analysis and Fisherâ□□s exact test (q-value

PrgmNr 3843 - Spatial Whole Transcriptome Analysis of human kidney histological Structures

View session detail

Author Block: **Y. Liang**[1], A. Vanshoiack[1], W. Yang[1], L. Zhang[1], P. Danaher[1], J. Reeves[1], E. Piazza[1], M. Grout[1], C. Glaser[1], D. Kroeppler[1], E. Brown[1], K. Ton[1], M. Hoang[1], G. Demirkan[1], K. Fuhrman[1], S. Ranade[1], M. D. Rhodes[2], S. Warren[1], J. M. Beechem[1]; [1]NanoString Technologies, Seattle, WA, [2]Nanostring Technologies, Seattle, WA
Disclosure Block: **Y. Liang:** Salary/Employment; NanoString Technologies.
A thorough understanding of normal tissue biology is crucial to advances in disease treatment. However, until the advent of spatial biology, it was challenging to generate detailed molecular profiles of the individual structures that comprise organ architecture and function. In this study, we report a deep understanding of kidney function by analyzing whole transcriptomes of histological structures, encompassing both what decades of detailed molecular studies have unveiled along with novel insights into organ physiology.
Using the GeoMx DigitalÂ® Spatial Profiler (DSP) and accompanying Whole Transcriptome Assay, we analyzed four non-diseased kidneys as a proof of principle for spatial Organ profiling studies. We profiled fundamental functional structures within kidney nephrons: glomeruli, glomerular filtration membrane, proximal and distal convoluted tubules, loops of Henle, and collecting ducts. GeoMxÂ® DSP allowed the selection of each of the histological structures for specific profiling of the whole transcriptome with high precision.
We detected 9,779 genes across the various structures, with the average detection in distinct compartments ranging from 3800-7000 genes. Data were benchmarked against the protein expression data from Human Protein Atlas as independent validation of key markers of distinct structures, with approximately 90% concordance between RNA expression and known protein localization. Additionally, deconvolution based on single-cell RNASeq enabled us to calculate the proportion of distinct cell types within each structure, confirming the previously known cell composition of these features. We also observed that each of the kidney sub-structures had enrichment in specific genes and pathways as per expectation, e.g., ion and metabolite transporters expressed explicitly in proximal and distal tubules, recapitulating 50 years of kidney physiology in a single experiment. Moreover, the retention of spatial information facilitated the resolution of differentially expressed genes between the cortical and juxtamedullary glomeruli Finally, the characterization of expression differences between the four kidney samples enables preliminary examination of the variation between individuals.
In conclusion, we have collected whole transcriptome data for establishing normal spatial gene expression profiles for key structural components of the kidney. These data will be publicly available and can be used as standards to inform future profiling studies in normal and diseased kidney tissue.

PrgmNr 3844 - Spatial whole transcriptome profiling uncovers unique functional insights into the histological structures of the human pancreas

View session detail

Author Block: **E. Piazza**[1], Y. Liang[1], A. Vanshoiak[1], W. Yang[1], L. Zhang[1], P. Danaher[1], J. Reeves[1], M. Grout[1], C. Glaser[1], D. Kroeppler[1], E. Brown[1], K. Ton[1], M. Hoang[1], G. Demirkin[1], K. Fuhrman[1], S. Ranade[1], M. D. Rhodes[2], S. Warren[1], J. M. Beechem[1]; [1]NanoString Technologies, Seattle, WA, [2]Nanostring Technologies, Seattle, WA

Disclosure Block: **E. Piazza:** Salary/Employment; NanoString Technologies.

Until the advent of spatial biology, it was not easy to generate detailed molecular profiles of the individual structures that comprise organ architecture and function. In this study, we report a deep characterization of pancreas function by analyzing whole transcriptomes of histological structures, encompassing both what decades of detailed molecular studies have unveiled along with novel insights into organ physiology.

Using the unique ability of GeoMxÂ® Digital Spatial Profiler (DSP) to select histological structures of organs specifically and measure their gene expression with the Whole Transcriptome Assay, we analyzed four non-diseased pancreas samples. Our data demonstrate capability of GeoMxÂ® to profile critical functional structures within the pancreas with high precision. We focused the study on the islet of Langerhans, acini, and ducts within the pancreas, quantitating expression of the whole transcriptome across these different structures with the average detection in distinct compartments ranging from 6,000-8,000 genes.

Data were benchmarked against Human Protein Atlas protein expression data as independent validation of key markers of distinct structures, with approximately 90% concordance between RNA expression and known protein localization. Deconvolution based on single-cell RNASeq enabled us to calculate the proportion of distinct cell types within each structure, confirming known cell compositions. For example, while ductal components consist of small islands of cells punctuating the acini, by direct illumination of these cells we identified over 3000 genes that were highly specific to either structure (FDR In conclusion, we have collected whole transcriptome data for establishing normal spatial gene expression profiles for key structural components of the pancreas. These data will be publicly available and can be used as standards to inform future profiling studies in normal and diseased tissue.

PrgmNr 3845 - Targeting clinically significant dark regions of the human genome with high-accuracy, long-read sequencing

View session detail

Author Block: **I. McLaughlin**, J. Harting, Z. Kronenberg, L. Aro, C. Heiner; Pacific BioSci.s, Menlo Park, CA
Disclosure Block: **I. McLaughlin:** Major Stockholder/Ownership Interest; Pacific Biosciences. Salary/Employment; Pacific Biosciences.
Introduction: There are many clinically important genes in â□□darkâ□□ regions of the human genome. These regions are characterized as dark due to a paucity of NGS coverage as a result of short-read sequencing or mapping difficulties. Low NGS sequencing yield can arise in these regions due to the presence of various repeat elements or biased base composition while inaccurate mapping is attributable to segmental duplications. Long-read sequencing coupled with an optimized, robust enrichment method has the potential to illuminate these dark regions. **Materials and Methods:** Using PacBio highly accurate long-read (HiFi) sequencing, coupled with a long-PCR targeted enrichment method, we investigated two important dark region genes that are challenging to accurately type with short-read sequencing due to associated pseudogenes: *CYP21A2*, responsible for congenital adrenal hyperplasia, and *GBA*, responsible for Gaucher disease. For each gene, our aim was to cover regions of pathogenic mutations in a single contiguous sequence or set of sequences that can be assayed in a single reaction. *CYP21A2* and an associated pseudogene *CYP21A1P* were co-amplified in a single long-range PCR reaction generating a 10.2 kb and 8.9 kb amplicon, respectively. Similarly, *GBA* and an associated pseudogene *GBAP1* were co-amplified in a single long-range PCR reaction generating a 12.6 kb and 16.0 kb amplicon, respectively. Seven Coriell samples for the *CYP21A2* target region and 13 Coriell samples for the *GBA* target region containing known pathogenic mutations were studied in replicate. SMRTbell libraries were generated from pooled amplicons for each target gene and sequenced on a PacBio Sequel II System. Accounting for replicates, each library contained a multiplex of 24 samples. A new PacBio sequence clustering algorithm, pbaa, designed for rapid analysis of HiFi reads from amplicons was used in variant typing. **Results:** All pathogenic *CYP21A2* and *GBA* variants were accurately called in the test samples. These variants included whole-gene deletions, gene duplication, gene fusions, recombinant exons, and phased compound heterozygotes. **Conclusion:** We demonstrate that HiFi sequencing provides new opportunities for sequencing clinically relevant but previously dark regions of the human genome that are underrepresented in short-read sequencing. Accurate long reads provide important phasing information, identify structural variations, and avoid potential confusion with pseudogenes. HiFi sequencing of these regions enables a better understanding of the relationship between genetic factors and personal health and has the potential to ultimately help guide health-related decisions.

PrgmNr 3846 - Automated End-to-End NGS Sample Preparation: DNA Extraction and Library Preparation on a Novel Digital Microfluidics System

View session detail

Author Block: **K. Cunningham**, E. Cervantes, J. Yoo, A. Barner, A. E. Carvajal, C-C. Lee, A. Chavan, A. Cervantes, S. Margeridon, M. Jebrail, F. Christodoulou; Miroculus, San Francisco, CA
Disclosure Block: **K. Cunningham:** None.
Whole genome sequencing (WGS) is increasingly used to detect variants in coding and non-coding regions of the human genome in clinical and research settings. The sample processing steps upstream of sequencing (including extraction of DNA and WGS library preparation) can be time-consuming, labor intensive, and subject to variations in benchtop technique. Automation solutions for processing whole blood are often standalone and disjointed from liquid handling instruments used for library preparation steps. Here, we introduce a novel digital microfluidic (DMF)-based system, composed of: the Miro Canvas instrument, single-use electronics-free cartridges, and an intuitive drag-and-drop protocol building application. Miro Canvas offers the convenience and consistency of automation, with the flexibility to run individual samples and easily implement new workflows. Miro Canvas has automated all the steps required for the Illumina Lysis Reagent Kit to extract DNA from whole blood, followed by the Illumina DNA PCR-Free Prep workflow. The steps include blood lysis, DNA binding to beads, washes and elution (cartridge 1), followed by tagmentation of purified DNA, bead cleanup, ligation, library cleanup and final elution (cartridge 2). Using 20 Î¼l of whole blood, we compared DNA extraction and library preparation automated on Miro Canvas to manual preparations. Final libraries were sequenced on the MiSeq platform (PE150). Automated libraries had comparable final library yields (9.67Â±1.56 nM vs. 10.73Â±1.58 nM), percent coverage (percent of genome with 1X coverage at 350,000-500,000 passed filter reads), and decreased percent excluded bases (an average of 0.9% less) when compared to manual libraries. In conclusion, the Miro Canvas system readily extracts gDNA from whole blood samples and creates consistent, high-quality NGS libraries for WGS with the hands-off automation of high throughput liquid handlers, and the flexibility to easily run individual samples on demand for a wide variety of protocols. This end-to-end workflow has the potential for personalizing and widely distributing WGS-based testing in clinical and research settings.

PrgmNr 3847 - Comparison of fully automated to Semi-Automated Variant Curation in Genetic Carrier Screening

View session detail

Author Block: **B. Gall**[1], T. Smart[2], S. Kolluri[1], H. Tadepally[1], K. Phaik Har Lim[1], N. Sanapareddy[1], D. Keen-Kim[1]; [1]Natera, Inc., San Carlos, CA, [2]Natera, Inc, San Carlos, CA
Disclosure Block: **B. Gall:** Major Stockholder/Ownership Interest; Natera, Inc.. Salary/Employment; Natera, Inc..
The rapid integration of artificial intelligence (AI) in genomic testing has resulted in the increased use of AI in genomic variant curation. While AI is highly efficient, it is hypothesized that manual curation can more accurately capture nuances such as â□□borderlineâ□□ or complex data in the literature. We therefore investigated internal data as a case study by comparing the sensitivity and specificity of one commercially-available fully automated variant curation software against an AI-assisted semi-automated curation using the same software. Though each process applies the ACMG guidelines to reach a final pathogenicity classification, insight into the AI adaptation of ACMG criteria was limited due to proprietary constraints of the automated variant classification software used for this comparison. We also compared the sensitivity and specificity of fully automated curations to curations where semi-automated curations and high-confidence ClinVar classifications were concordant. We performed our analysis using historical data from the HorizonTM carrier screening panel (Natera, Inc.), collected between November 2019 and November 2020. Our analysis of fully-automated variant curation compared to semi-automated variant curation in calling a variant positive (likely pathogenic or pathogenic) or negative (variant of uncertain significance (VUS), likely benign and benign) had a sensitivity and specificity of 90.60% and 97.39%, respectively (n=116,721). Comparison of a fully-automated variant curation and a semi-automated curation which was concordant with high-confidence ClinVar classifications resulted in a sensitivity and specificity of 77.29% and 98.28%, respectively (n=5,531). The likelihood of a false positive report from a fully-automated classification was calculated to be between 9.40-22.71%. This disparity is likely a result of inadequate AI collation and interpretation of literature-derived data necessary for applying ACMG criteria. Analysis of a subset of these discrepancies (n=78) occurring over a period of six months confirmed that AI-computed curations omitted or overinterpreted literature with clinical cases (24.36%) and/or failed to interpret mixed pathogenic and benign evidence (15.38%). Despite the rising popularity of AI, our data comparing one AI curation program to semi-automated classification demonstrates that AI-computed curation is less accurate than semi-automated classification. Our analysis supports the use of manual curation in combination with AI-supported curation, to increase productivity and streamline processes, while preserving the accuracy associated with manual curation.

PrgmNr 3848 - Designing Probesets for Interfering Variant Awareness on Genotyping Microarrays

View session detail

Author Block: **J. Fang**, A. Mittal, J. Gollub, K. Aull, A. H. Roter, J. P. Schmidt; Thermo Fisher Scientific, Santa Clara, CA
Disclosure Block: **J. Fang:** Salary/Employment; Thermo Fisher Scientific.
The number of known polymorphic variants in the human genome has increased significantly throughout the past decade, ranging from about 80 million variants in the 1000 genomes Phase 3 study (2015) up to over 700 million variants in the gnomADv3.1 study (2020). A significant fraction of these polymorphisms overlap in genomic locus, resulting in potentially multiallelic loci. However, public databases of human polymorphisms such as dbSNP and gnomAD do not consistently account for these possible overlaps. To accurately genotype a target locus on the hybridization-based genotyping microarrays, it is crucial to account for interfering polymorphisms. Failure to do so may result in unintended hybridization to spurious targets. Using the gnomADv3.1 data set, we performed a comprehensive check for overlapping variants against several different catalogs of known mutations, while also accounting for varying allele frequencies across different superpopulations. The ClinVar database of ~625,000 variants (Dec 2020) are particularly enriched in overlapping variants. By screening the ClinVar variants against the 707.9 million known polymorphic sites in gnomADv3.1 across 5 superpopulations (AFR, AMR, EAS, NFE, SAS), we found that 31.7% of the Clinvar variants overlapped with at least one other polymorphic site. Within the ACMG 59 genes, 30% of ClinVar variants overlapped another polymorphic site. We also demonstrate the ability to correctly genotype overlapping variants by using probes specifically designed to factor in the overlapping variants on the Applied Biosystemsâ□¢ CarrierScanâ□¢ and AxiomTM Precision Medicine Diversity (PMD) Research arrays. We show several examples of pathogenic variants associated with severe traits and conditions where multiallelic probe designs were able to correctly resolve the more common interfering variant from the relatively rare target variant. Example of overlapping variants includes the short deletion rs1805177 overlapping with the SNP rs10229820 with MAFs of 6.2% and 13.1% respectively, and the rare delATC variant rs121908745 overlapping with the relatively common delCTT variant rs113993960 with MAFs of 0.014% and 0.26% respectively (AFR superpopulation, gnomADv3.1). Both complex loci were resolved using multiallelic genotyping on the CarrierScanâ□¢ array.This study demonstrates the importance of screening and designing for overlapping variants. With the steadily increasing number of rare genomic variants being published in literature, common interfering variants become increasingly relevant.

PrgmNr 3849 - DNAscopeTM: A novel chromogenic in-situ hybridization technology for high-resolution detection of DNA copy number and structural variations

View session detail

Author Block: **V. Murlidhar**, L-C. Wang, F. Tondnevis, C. Todorov, J. Gaspar, A. Sahajan, B. Zhang, X-J. Ma; Bio-Techne, Newark, CA
Disclosure Block: **V. Murlidhar:** Salary/Employment; Bio-Techne.
Genomic DNA anomalies such as copy number variations (gene duplication, amplification, deletion) and gene rearrangements are important biomarkers and drug targets in many cancer types. DNA in-situ hybridization (ISH) is the gold standard method to directly visualize these molecular alterations in formalin-fixed paraffin-embedded (FFPE) tumor tissues at single-cell resolution within a histological section. However, currently available fluorescent ISH (FISH) assays provide limited morphological detail due to the use of fluorescent nuclear staining compared to chromogenic staining. Furthermore, FISH techniques rely on expensive fluorescence microscopes, risk loss of fluorescent signal over time and involve tedious imaging at high magnifications (100X). There is thus an unmet need for a sensitive and robust chromogenic DNA-ISH assay that can enable high-resolution detection of genomic DNA targets with the ease of bright-field microscopy. We present here DNAscope - a novel chromogenic DNA-ISH assay - for detecting and visualizing genomic DNA targets under a standard light microscope. DNAscope is based on the widely used RNAscopeÂ® double-Z probe design and signal amplification technology and provides unparalleled sensitivity and specificity with large signal dots readily visualized at 40X magnification and with full morphological context. Furthermore, DNAscope ensures specific DNA detection without interference from RNA due to the use of a novel RNA removal method. Using a duplex chromogenic detection assay in red and blue, we demonstrate highly specific and efficient detection of gene rearrangements (ALK), gene amplification (ERBB2, EGFR, MET) and deletion (TP53 and CDKN2A). The DNAscope assay has been carefully optimized for probe signal size and color contrast to enable easy interpretation of signals under conventional light microscopy or digital pathology. Compared to conventional FISH assays, DNAscope probes are standard oligos that are designed in silico to be free of any repetitive sequences and can be rapidly synthesized for any DNA target. In conclusion, the DNAscope assay provides a powerful and convenient alternative to commonly used FISH assays in many research applications.

PrgmNr 3850 - Immunophenotyping of TCR and BCR clonotypes

View session detail

Author Block: **A. Chenchik**, M. Makhanov, R. Darst, T. Liu, P. Diehl, L. Kobzik; Cellecta, Mountain View, CA
Disclosure Block: **A. Chenchik:** None.
TCR and BCR repertoire profiling holds great potential for understanding the disease mechanisms and development of new therapeutics in infectious disease, auto-immunity and immuno-oncology. However, this potential could be greatly improved by combining information about receptor clonotypes with immuno-phenotypes of T and B cells. To facilitate these studies, we developed a novel technology for combined profiling of all human TCR and BCR variable regions and phenotypic characterization of immune cells in bulk and at the single-cell level in PBMC and immune cell fraction samples. The developed TCR/BCR Immunophenotyping method involves multiplex RT-PCR amplification and sequencing of CDR regions of TCR and BCR genes and of a set of the most informative T- and B-cell phenotyping genes. Bioinformatics analysis of NGS data allows us to profile TCR/BCR clonotypes, as well as identify major immune cell subtypes and their activation status. Data will be presented showing how combined TCR/BCR clonotype analysis combined with targeted expression profiling of immune cells can be applied for large-scale discovery of novel cell typing and activation biomarkers in several immune-response model systems. Preliminary studies demonstrate the assay has unparalleled throughput, sensitivity, and improved cost-effectiveness for high-throughput immunity biomarker discovery applications.

PrgmNr 3851 - Implications of biological networks: Search for the basal gene network using plasma proteomic signatures of COVID-19 patients

View session detail

Author Block: **A. Podder**[1], N. J. Schork[1,2]; [1]Translational Genomics Res. Inst. (TGen), phoenix, AZ, [2]The City of Hope Natl. Med. Ctr., Duarte, CA
Disclosure Block: **A. Podder:** None.

The prime components of human biological machinery, e.g., genes, proteins, and metabolites, do not function in isolation. Instead, a range of specific interactions occur between various micro- and macro-molecules inside a cell. These molecular interactions define nearly all biological processes, and their knowledge is essential to manage the perturbations contribute to the onset of disease pathogenesis. Such interactions can be outlined by the graph theory, which uses assorted models from mathematics and physics. In the past decade, significant progress has been made in researching common diseases, and numerous high-throughput datasets have emerged across multiple layers of cellular physiology including genomic, transcriptomic, proteomic, and metabolomic. These datasets can be utilized to construct a graph-based representation of biological interactions to delineate critical gene regulatory processes, protein-protein interactions, signal transduction pathways, metabolites, and reaction maps that can be interrogated to understand disease processes.

We have investigated such elemental molecular interactions for various diseases and traits, including COVID-19, the current pandemic that has spread rapidly across the globe. Even after rigorous year-long research, COVID-19 poses a significant risk to human health, we therefore have actively sought to understand how the SARS-CoV-2 virus of the Corona-virus family infects the human body and what molecular disturbances take place during infection. We employed a network approach based on physical protein-protein interactions to find cellular processes dysregulated as a whole in response to a COVID-19 infection. For this, we have used available data reflecting the differential expression of 1,420 plasma proteins for 306 COVID-19 patients and 78 control individuals featured across four different time points in a three weeks infection. We show that, based upon this approach, we can uncover novel insights and promising drug targets for COVID-19 that go beyond those focusing on limited list of individually altered genes.

PrgmNr 3852 - Large-scale functional assays to comprehensively assess SNP-accessible variants in the Urea Cycle Disorder genes, *OTC*, *ASS1*, and *ASL*

View session detail

Author Block: **G. Cromie**[1], R. Lo[1], K. Owens[2], A. Sirr[1], J. N. Kutz[2], R. McLaughlin[1], A. M. Dudley[1]; [1]Pacific Northwest Res. Inst., Seattle, WA, [2]Dept. of Applied Mathematics, Univ. of Washington, Seattle, WA

Disclosure Block: **G. Cromie:** None.

Here, we report our progress in developing high-throughput functional assays of variant effect for the three human urea cycle disorder (UCD) genes encoding ornithine transcarbamylase (OTC), argininosuccinic acid synthetase (ASS1), and argininosuccinic acid lyase (ASL).

Urea cycle disorders are highly actionable diseases for which understanding the functional implications of genetic variation in the causative genes can positively impact patient health. Neonatal onset is associated with severe enzyme deficiency. These infants appear normal at birth, but can rapidly accumulate high levels of ammonia, potentially resulting in cerebral edema, lethargy, seizures, coma, and death. As such, in the United States, biochemical assays of ASL and ASS1 protein function are included in the newborn screening panels of all 50 states, and OTC activity is screened in eight. In contrast, individuals harboring moderate to mild loss of function alleles may remain undiagnosed into childhood or even adulthood. Late onset UCDs generally involve environmental triggers (e.g. surgery, pregnancy, or chemical exposure) in individuals with reduced enzyme function. The late-onset form often presents with episodic psychosis, bipolar disorder and major depression, and without treatment, prognosis is poor. Thus, knowledge of the functional implications of genetic variation in these genes could reduce the morbidity and mortality associated with delayed treatment or underdiagnosis.

Our approach leverages the advent of low-cost, large-scale gene synthesis and the development of high-throughput *in vivo* assays of protein function (Multiplexed Assays of Variant Effect, MAVEs) in the model organism, *Saccharomyces cerevisiae*. Because OTC, ASS1 and ASL are enzymes in the highly conserved arginine biosynthesis pathway, the human protein coding sequences can functionally replace (genetically complement) deletions of their corresponding yeast orthologs. Therefore, yeast growth in the absence of arginine provides a quantitative readout of human enzyme function with disease alleles showing significant reduction relative to wild-type. We will present quantitative results for the functional impact of 1755 (93%) of all SNP-accessible amino acid substitutions in the OTC protein. We also discuss these results in the context of the published OTC crystal structure and the extent to which they agree or conflict with published annotations of pathogenicity, and when available, disease severity.

PrgmNr 3853 - Optimizing performance of copy number variant calling based on improved baseline selection

View session detail

Author Block: **A. Bhattacharya**, T. Angeloni, D. Goos-Root, D. Xu, A. Smolgovsky, J. Wright, K. Kazmierkiewicz, A. Grupe, R. Owen, E. Moler, S. Daniel, F. Lacbawan; Quest Diagnostics, San Juan Capistrano, CA
Disclosure Block: **A. Bhattacharya:** None.
Detecting copy number variation (CNV) using read count from next-generation sequencing (NGS) data is often challenged by separating true CNV events from background noise. Commonly used read count-normalization techniques address this challenge by removing various sources of background noise (eg, batch effects, GC bias, or low coverage in low complexity or pseudogene regions) or by removing higher order principal components. Even after normalization, CNV callers often fail to separate real events when coverage is below 200X or when event sizes are small or partially covered by capture probes. To address the challenges, we developed a computational approach that uses a systematic selection of baseline (normal) samples. This new baseline-selection method initially includes all the samples in a sequencing batch as initial baseline and then excludes all the low coverage and volatile samples to form final baseline.
Detection rates were compared for our CNV calling pipeline (detecting outliers from normalized coverage data) with the new versus old (baseline includes all samples in a sequencing batch) baseline-selection methods. Coverage data from 43 low-coverage (average coverage less than 200x) samples with 118 confirmed CNV events (using a validated orthogonal method i.e., array, MLPA) were used; 40% of CNVs were small 1- to 2-exon deletions or duplications. Without the baseline-selection, 39 (33%) CNVs were detected. With baseline-selection, 112 (95%) CNVs were detected. These results show the baseline-selection method presented here can improve detection of CNVs in samples with low coverage.

PrgmNr 3854 - Plasma protein profiling of Alzheimerâ□□s and mild cognitive impairment subjects with a novel approach for identification of known and unknown candidate biomarkers

View session detail

Author Block: **R. Benz**, M. Goldberg, J. C. Cuevas, W. Manning, X. Zhao, T. L. Platt, M. Ko, H. J. Ha, H. Liou, E. M. Elgierari, M. Figa, H. Xia, D. Harris, P. Ma, O. C. Farokhzad, J. E. Blume, A. Siddiqui; Seer, Redwood City, CA
Disclosure Block: **R. Benz:** Salary/Employment; Seer.
Blood plasma, which contains valuable information from most tissues, is a rich source of protein biomarkers for early detection of diseases, but its large dynamic range of protein concentrations necessitates complex workflows and trade-offs between throughput, scalability, coverage, and precision. Here we use a deep and quantitative proteome profiling platform, Proteograph™ Product Suite, which leverages multiple nanoparticles, engineered with distinct physicochemical properties to provide broad coverage of the plasma proteome at scale. In this study we aim to identify protein biomarkers for Alzheimerâ□□s disease (AD) from blood with this untargeted plasma protein profiling approach, for AD and Mild Cognitive Impairment (MCI) conditions.
Methods. Plasma samples from 200 subjects comprising 50 AD, 50 MCI, and 100 controls were profiled using the Proteograph plasma protein profiling platform[1]. Using a 5-nanoparticle panel and 85 ÂµL of plasma per nanoparticle, proteins were quantified by data-independent acquisition (DIA) liquid-chromatography mass-spectrometry (LC-MS) in about 6 weeks. Normalized peptide intensities were used in ten rounds of 10-fold cross-validation to develop random forest models for class discrimination.
Results. Across the samples, 2,391 plasma proteins were detected, with 2,085 in at least 25% of samples. 36 proteins with the highest AD OpenTargets[2] scores were detected, including Amyloid Beta (A4) Precursor Protein. 25,593 protein-comprising peptides were detected, with 15,661 in at least 25% of the samples. Univariate analysis identified 441 and 526 proteins that were significantly different in AD or MCI versus control, respectively. Random-forest classification for AD and MCI produced ROC AUCs that were at least 0.90. Top features by importance included known and unknown candidate biomarkers.
Conclusions. These analyses have identified novel combinations of candidate plasma protein markers, many without prior known relevance to AD. More broadly, the Proteograph platform confirmed its ability to generate profiling data in a deep, broad, and rapid fashion, enabling large-scale studies to detect novel insights with clinically useful potential.
References.
1. Blume JE, et al., Nat. Comm. 2020;11(1):3662-1
2. https://www.opentargets.org

PrgmNr 3855 - Proteogenomic analyses of non-small cell lung cancer subjects identifies candidate lung-cancer associated protein isoforms and protein variants

View session detail

Author Block: **M. Donovan**, J. Blume, M. Ko, R. Benz, T. Platt, J. Cuevas, O. Farokhzad, S. Batzoglou, A. Siddiqui; Seer Bio, Redwood City, CA
Disclosure Block: **M. Donovan:** None.
Introduction: Next-generation sequencing studies at scale have revealed novel biological insights into the molecular link between genotype and human biology through the measurement of the multiple biological steps occurring between genome, epigenome, and transcriptome. Similarly scaled and comprehensive studies examining the human proteome have been challenging to achieve due to the complexity of distinguishing between multiple protein variants arising from a single gene due to allelic variation, alternative splicing, post translational modifications, and degradation. Scalable, deep and unbiased proteomics studies have been historically impractical due to cumbersome and lengthy workflows required for complex but easily accessible samples, like blood plasma. Here, we demonstrate the power of Proteograph, a next generation nanoparticle-based automated plasma proteome profiler, in a proof-of-concept proteogenomic analysis of 80 healthy controls and 61 early stage non-small-cell lung cancer (NSCLC) samples to dissect differences between protein isoforms arising from alternative gene splicing and to identify peptide variants arising from allelic variation.
Methods and Results: Processing the 141 plasma samples with Proteograph yielded 21,959 peptides corresponding to 2,499 protein groups. Using peptides with significant abundance differences (p **Conclusions**: Proteograph can generate unbiased and deep plasma proteome profiles that enable identification of protein isoforms and variants present in human plasma with potential to scale to sufficient sample sizes enabling population-scale proteogenomic studies.

PrgmNr 3856 - Proteograph™ Analysis Suite: A cloud-scalable software suite for proteogenomics data analysis and visualization

View session detail

Author Block: **A. Vadapalli**, Y. Berk, H. Auluck, A. Gajadhar, Y. Lou, A. Siddiqui; Seer, Inc, Redwood City, CA
Disclosure Block: **A. Vadapalli:** Salary/Employment; Seer, Inc.
Researchers are increasingly adopting multi-omics approaches to understand the complex biological processes that underlie human diseases. Next generation sequencing (NGS) is widely used for identifying genetic variants and gene function while mass-spectrometry is used to quantify protein abundances, modifications, and interactions from complex samples like plasma. A new plasma profiling platform called the Proteograph Product Suite was developed that leverages multiple nanoparticles with distinct physiochemical properties to provide deep plasma proteomic analysis at scale. The analysis of proteomics and genomics data typically requires a wide collection of different tools, which is further complicated by the prevalence of command-line interfaces and operating system-specific requirements that can act as a barrier for researchers to adapt new data analysis tools due to their steep learning curve and implementation costs. In this abstract, we present a cloud-based analysis software platform called Proteograph Analysis Suite (PAS) that analyzes proteomics data derived from the Proteograph workflow along with genomic variant results imported from NGS experiments. The main features of the software suite include an experiment data management system, analysis protocols, an analysis setup wizard, and tools for reviewing and visualizing results. PAS can support both Data Independent Analysis (DIA) and Data Dependent Analysis (DDA) proteomics workflows and is compatible with widely accepted format of variant call files from NGS workflows. For each analysis run, users are able to view various quality control metrics like peptide/protein group intensity, protein sequence coverage, relative protein abundance distribution, peptide and protein groups stratified by nanoparticle. Various visualizations such as principal component analysis, hierarchical clustering, and heatmaps allow intuitive identification of dataset trends. Quantitative differential expression tools such as volcano plots, protein interaction maps and protein-set enrichment simplify data interpretation enabling functional insights.In conclusion, we present a comprehensive proteogenomic analysis software suite to enable user-friendly and reproducible multi-omics analyses of proteomic and genomic data.

PrgmNr 3857 - Quantification of gene expression changes in mouse disease models using a high-throughput spatial omics platform

View session detail

Author Block: **T. Awad**, O. Marcos, R. Yin, N. Kotova, I. Oh; Rebus BioSci.s, Santa Clara, CA
Disclosure Block: **T. Awad:** None.
Batch effects due to technical variability are a major problem in single cell transcriptomics. Spatial methods are no exception - their low throughput requires high numbers of technical replicates, reducing the statistical power needed to quantify differential gene expression across experimental conditions. To overcome this problem, we took advantage of the large imaging area of the Rebus Esper spatial omics platform. We processed brain sections from three mouse genotypes in parallel - one wild type and two disease models. In addition, the Esper High Fidelity assay, based on single-molecular fluorescent in situ hybridization (smFISH), requires no amplification, yielding quantitative results with minimal batch effect. The combination of low technical variation and balanced experimental design allowed us to integrate more than 500,000 cells from multiple datasets for analysis without the need for batch correction. We were able to identify more than 12 neuronal and glial cell type clusters using 20 cell type-specific genes, and further dissect these cell types by anatomical structure utilizing the spatial information. We then performed differential gene expression of 10 disease-related genes on each cell-type subset. These results demonstrate the ability of the Rebus Esper spatial omics platform to yield high-throughput spatial omics data with single cell resolution, low technical variation, and the sensitivity and specificity required for differential gene expression quantification.

PrgmNr 3858 - Reducing variability of breast cancer subtype predictors by grounding deep learning models in prior knowledge

View session detail

Author Block: **J. Davidson**, W. A. Johnson, R. Gadgil, E. Schwab, P. Anderson; Cal Poly, San Luis Obispo, San Luis Obispo, CA
Disclosure Block: **J. Davidson:** None.
INTRODUCTION:The benefit of biomedical informatics research is the synthesis of computational power and biomedical insight. Deep learning networks in biology exemplify such a synthesis, having successfully streamlined various bioinformatics problems like cancer subtype classification, drug sensitivity prediction, and imaging analysis. However, many networks suffer from underspecification, with multiple groups of near-optimal predictors. This is quite problematic in biomedicine, where datasets increasingly include far more input predictors than data samples. Deep neural networks allow for analysis of such high-dimensional data without pre-processing or feature extraction. We aim to provide a framework in which curated knowledge can be added to a neural network to address underspecification and improve network interpretability. METHODS: We developed a three-stage pipeline for incorporating prior knowledge into predictive modeling. Our approach encourages researchers to share and reuse curated knowledge and methodology in an iterative manner. For this study a 3-gene model incorporates master regulator genes ESR1, ERBB2, and AUKRA as appropriate predictors are for subtypes basal-like, HER2-enriched, luminal A, and luminal B. We combine curated knowledge with an experimental dataset to produce derived knowledge, with the goal of quantifying the derived knowledge for downstream modeling and analysis. For this study, our experimental data are gene expression values from a set of 2,133 participants of the METABRIC cohort with known subtype classifications. The patient data and the curated knowledge are combined to produce a graph of patient-patient relationships that encodes which patients exhibit similar profiles under the three-gene model. The final stage, derived knowledge from stage 2 are embedded into machine learning systems. In this work, we apply the derived knowledge from stage 2 to the deep learning neural network designed for breast cancer subtyping. RESULTS:Our approach introduces a new parameter, knowledge loss weight, to study the effect of increasing the importance of knowledge in the overall loss function. We studied the performance of the model after the incorporation of knowledge into the loss function both on the training and the validation dataset. We studied the stability of the results by examining the variability of the importance of each gene over multiple repetitions of the training algorithm. CONCLUSIONS: Combining the versatility and predictive power of deep neural networks with the insight of biological knowledge provides a solution to the problems of data scarcity and model interpretability in bioinformatics.

PrgmNr 3859 - 'Drugging' PTPRD (AD, RLS and Addiction-associated gene) PTPRD

View session detail

Author Block: **G. R. Uhl**[1], W. Wang[2], T. Prisinzano[3], D. Johnson[4], I. Henderson[1]; [1]UNM/NMVAHCS, Albuquerque, NM, [2]Univ. of Arizona, Tucson, AZ, [3]Univ. of Kentucky, Lexington, KY, [4]Univ. of Kansas, Lawrence, KS

Disclosure Block: **G.R. Uhl:** None.

Variants in the human PTPRD gene provide polygenic associations with vulnerability to addictions and ability to quit. They provide oligogenic associations with densities of neurofibrillary pathology (NFTs) in Alzheimer's disease (AD) brains, vulnerability to restless leg syndrome (RLS) and levels of PTPRD mRNA expression in postmortem human brains. Mouse models support the findings in addiction and RLS. In vitro data supports association with AD NFTs since PTPRD dephosphorylates the regulatory phosphorylation of GSK3s, principal serine/ threonine kinases on the path to forming NFTs. In aggregate, these data support strategies to reduce PTPRD phosphatase activity to reduce reward from addictive substances and to enhance this activity in order to reduce vulnerability to AD NFTs and possibly reduce symptoms in RLS.We have synthesized and tested over 40 novel analogs of our lead compound PTPRD phosphatase inhibitor, 7-BIA, and have identified an analog with greater potency, selectivity vs other phosphatases, selectivity in relation to targets of other known drugs, oral bioavailability, therapeutic index between 100 and 1000x proposed therapeutic dose and ability to reduce a measure of stimulant reward, cocaine conditioned place preference, by half in wildtype mice but not in mice with reduced PTPRD expression. We have also identified flavonol positive allosteric modulation of PTPRD's phosphatase. Quercetin can enhance PTPRD's ability to dephosphorylate GSK3 by more than 1.5-fold. Our results support the ability of a target identified by oligo- and polygenic association signals to provide a site for development of potentially useful novel therapeutics. Quercetin is generally recognized as safe by the FDA, facilitating human trials of its tolerability and utility in slowing progression to AD in aging.

PrgmNr 3860 - Development and optimization of a 43-gene pharmacogenomic panel using enrichment-based capture and PacBio HiFi sequencing

View session detail

Author Block: **D. Portik**[1], T. Hon[1], J. Wilcots[1], Y. Yang[2], N. A. Hammond[3], Z. Kronenberg[1], N. Watson[3], J. Harting[1], E. Ashley[4], J. Ziegle[1], S. A. Scott[5], S. B. Kingan[1]; [1]Pacific BioSci.s, Menlo Park, CA, [2]Stanford Univ. Sch. of Med., Stanford, CA, [3]Clinical Genomics Lab., Stanford Hlth.Care, Palo Alto, CA, [4]Stanford Univ., Stanford, CA, [5]Stanford Univ., Palo Alto, CA
Disclosure Block: **D. Portik:** Salary/Employment; Pacific Biosciences.
Pharmacogenomic testing in personalized medicine has the potential to improve patient treatment outcomes and reduce healthcare costs associated with medication efficacy and adverse drug reactions. Several technologies are commonly used for germline pharmacogenomic testing, including targeted genotyping, short-read sequencing, hybridization arrays, qPCR, and MLPA. Unfortunately, many clinically-significant pharmacogenomic loci remain challenging to accurately interrogate by these methods due to low sequence complexity and/or the presence of highly homologous pseudogenes. Long-read amplicon sequencing using Pacific Biosciences (PacBio) technology has previously been reported to accurately and precisely interrogate problematic pharmacogenomic loci, including *HLA*, *CYP2D6*, and *SLC6A4*; however, multi-gene pharmacogenomic HiFi sequencing panels have not been described. We developed a novel method to comprehensively interrogate a panel of 43 pharmacogenomic genes using an enrichment-based capture strategy (IDT) coupled with HiFi sequencing. Gene selection was centered on incorporating content with Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines and/or FDA recommendations, including selected variants with evidence for clinical actionability, translating to 935 kb of enrichment content. Performance of the novel panel and sequencing strategy was evaluated by subjecting reference material specimens with previously characterized variant calls and star (*) allele diplotypes to enrichment and HiFi sequencing on the Sequel IIe System. Consensus HiFi read clustering and alignment to GRCh38 used pbaa and pbmm2, respectively. Small variants (CYP2D6 variants, and enrichment-based HiFi sequencing correctly called the *CYP2D6*5* deletion allele in HG00276. These data indicate enrichment-based capture and HiFi sequencing is an effective approach for multi-gene panels, including historically challenging pharmacogenomic genes, for accurate variant discovery and full-gene haplotype phasing.

PrgmNr 3861 - Genomic analysis of angiotensin converting enzyme (ACE) inhibitor - induced angioedema

View session detail

Author Block: **M. A. Kennedy**[1], Y. Liau[1], S. D. S. Maggo[1], K. Lehnert[2], M. P. Doogue[1], A. Scull[1], A. L. Miller[1], J. F. Pearson[3]; [1]Univ. of Otago, Christchurch, New Zealand, [2]Univ. of Auckland, Auckland, New Zealand, [3]Univ. of Otago Christchurch, Christchurch, New Zealand
Disclosure Block: **M.A. Kennedy:** None.
Angioedema is a rare but serious adverse drug reaction associated with angiotensin converting enzyme inhibitors (ACEi). Prior research has indicated that specific ethnic populations are at a higher risk of ACEi - induced angioedema (ACEi-A), and several hereditary forms of angioedema exist, suggesting a genetic link. Several candidate gene and genome wide association studies have been conducted for ACEi-A, but no candidate genes or genetic variants of major effect have been identified. Using a model of monogenic predisposition to this adverse reaction, we set out to find variants that were rare in the population but highly enriched in ACEi-A. To achieve this, we conducted whole exome sequencing or whole genome sequencing on 20 ACEi-A cases. We identified one synonymous exonic variant in the *ACE* gene (rs4365) that was present in 40% of our ACEi-A cases compared to 4% of our controls (p-value=1.7e-05) and 3% of the gnomAD total dataset (p-value ACE and other genes. Using the WGS data, we also sought variants in candidate genes previously implicated in drug induced or hereditary angioedema, but found no evidence for consistent involvement of such variants in our cohort. In a separate analysis on the same dataset, we applied SKAT-O to evaluate rare variants within a larger set of genes linked to angioedema, and the main finding here was that aggregated variants in *ACE2* were nominally associated with reduced risk of angioedema. Finally, we examined copy number variants (CNV) called from the WGS data, and identified several rare duplication/deletion events in ACEi-A cases. A duplication encompassing *ADGRE1* was present in three angioedema cases, although this CNV is rare in the wider population. In conclusion, this analysis identified several interesting potential genetic contributors to ACEi-A, however, the small cohort size limits statistical power and these findings need to be evaluated in larger cohorts.

PrgmNr 3862 - Identification of pharmacogenetic predictors of methotrexate-induced mucositis in children with cancer

View session detail

Author Block: **C. Loucks**[1,2], A. Man[1,2], K. Li[1,2], S. R. Rassekh[1,3], C. J. Ross[4,1,5], B. C. Carleton[1,2,5,6]; [1]BC Children's Hosp. Res. Inst., Vancouver, BC, Canada, [2]Div. of Translational Therapeutics, Dept. of Pediatrics, Faculty of Med., Univ. of British Columbia, Vancouver, BC, Canada, [3]Div. of Oncology, Hematology and Bone Marrow Transplant, British Columbia Children's Hosp. and Univ. of British Columbia, Vancouver, BC, Canada, [4]Univ British Columbia, Vancouver, BC, Canada, [5]Dept. of Med. Genetics, Faculty of Med., Univ. of British Columbia, Vancouver, BC, Canada, [6]Pharmaceutical Outcomes Programme (POPi), British Columbia Children's Hosp., Vancouver, BC, Canada
Disclosure Block: **C. Loucks:** None.

Methotrexate is effectively used to treat acute lymphoblastic leukemia (ALL), yet some patients develop mucositis, a painful inflammation/ulceration of mucosal linings found at various locations throughout the body (e.g., the eyes, ears, mouth, digestive tract and vagina). In severe cases, mucositis is life-threatening due to increased susceptibility to infection or the inability to proceed with anti-cancer drug treatment. Studies exploring the potential impact of individual genetic susceptibility to methotrexate-induced mucositis have identified 28 genetic variants; however, their predictive value remains uncertain due to limited replication attempts or inconsistency between studies. Here, we aimed to discover and replicate genetic variants predictive of methotrexate-induced mucositis. We recruited children with ALL from six Canadian pediatric oncology centres and collected genomic DNA, captured detailed clinical/demographic information and conducted rigorous characterization of mucositis events to determine the probability they were methotrexate-induced. Overall, 584 methotrexate-treated children were enrolled in this study, and an observational case-control study design using logistic regression is being used to identify genetic predictors of mucositis. Cases (n=130; 22.2%) suffered serious mucositis defined as CTCAE/WHO grade â¥ 2 toxicity, controls (n=366; grade 0) did not experience mucositis, and children with mild mucositis (n=88; grade 1) were excluded to improve case/control separation. In preliminary genetic analyses, we first focused on children treated with high-dose methotrexate (â¥ 1000 mg/m^2; 86 cases; 192 controls), in which the majority of previous pharmacogenetic studies have been conducted. While these analyses revealed no significant associations with previously-identified variants (P â¥ 0.05), we uncovered an association with a novel genetic variant in close proximity to two genes (*AMDHD1* and *HAL*) involved in histidine catabolism (recently linked to methotrexate sensitivity) that puts patients at 5 times increased risk of developing serious mucositis (95% CI: 2.4-10.2; P=2.4x10^{-5}). Analyses are underway to determine the consistency/robustness of genetic findings by assessing the impact of case definition (e.g., considering more serious cases), treatment regimen (i.e., by incorporating children treated with lower methotrexate doses), and clinical factors (e.g., by adjusting analyses for radiation treatment that also impacts mucositis). Ultimately, identified genetic variants will form the basis of a mucositis-predictive test to help select the safest and most effective treatment strategy.

PrgmNr 3863 - Latino and Indigenous American populations may have five times as many pharmacogenetic variants as Europeans yet remain underrepresented in pharmacogenetic studies

View session detail

Author Block: **S. Hernandez**[1], L. A. Hindorff[2], J. Morales[1], E. M. Ramos[2], T. A. Manolio[3]; [1]Natl. Human Genome Res. Inst., Bethesda, MD, [2]NIH, Bethesda, MD, [3]NHGRI, Bethesda, MD
Disclosure Block: **S. Hernandez:** None.
Background: Pharmacogenetic (PGx) variants play a key role in the safety and efficacy of commonly prescribed drugs and are known to differ across ancestral groups, but frequency and impact of PGx variants among persons of Latino and Indigenous American ancestry are less well-characterized than in persons of European ancestry. If unrecognized, these differences may worsen health disparities experienced by Latino and Indigenous American populations. **Methods:** Allele frequencies for altered function variants in 11 PGx genes in â□□European,â□□ â□□Latino,â□□ and â□□Americanâ□□ populations were extracted from published Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines. CPIC reviews and curates published allele frequencies to calculate population allele frequencies for each gene, using PharmGKB definitions for European (primarily European descent, including European Americans), Latino (Mestizo descent, from Latin America, and self-identified Latino persons in the US), and American (from the Americas with ancestors predating European colonization) ancestries. After removing obvious duplicate persons reported in large databases such as gnomAD, total numbers of persons genotyped in CPIC guideline data for each of 11 genes with variants classified as normal or altered function (*CACNA1S, CYP2B6, CYP2C9, CYP2C19, CYP2D6, CYP3A5, DPYD, NUDT15, SLCO1B1, TPMT, UGT1A1*) were summed to calculate percentages of representation of these three groups and compare allele frequencies. **Results:** Allele frequencies for these 11 genes were reported for 632,000 persons of European, Latino, or American ancestry, of whom 520,000 (82%) were of European, 99,000 (16%) of Latino, and 13,000 (2%) of American ancestry. In contrast, population sampling consistent with 2019 US Census estimates would have yielded proportions of 75%, 23%, and 1.3%. Of 78 alleles with altered function, 28 were detected only in Latino or American populations or at $\geq$ 2-fold higher frequency than in European populations, while 27 alleles were detected only or at $\geq$ 2-fold higher frequency in European than Latino and American populations. Yield of altered function alleles were thus 0.52/100,000 European persons genotyped compared to 2.5/100,000 Latino or American persons genotyped. **Conclusions:** Allelic data used in CPIC guidelines for 11 PGx genes underrepresent Latino populations, while Latino and Indigenous American populations combined have a roughly 5-fold higher frequency of altered function variants. This could have important clinical implications and widen health disparities experienced by these two groups as PGx recommendations are brought into clinical practice.

PrgmNr 3864 - Research clinical pilot project for in silico pharmacogenomics concurrent with whole-genome sequencing in the pediatric ICU

View session detail

Author Block: **S. Batalov**, A. Hemperly, K. Rodriguez, A. Besterman, E. Sanford Kobayashi, N. Coufal, D. Dimmock, M. N. Bainbridge; Rady Children's Hosp., San Diego, CA
Disclosure Block: **S. Batalov:** None.
Pharmacogenomics is an important field for precision medicine, however its utility in the Pediatric and Neonatal intensive care units is not known. Rady Childrenâ□□s Clinical Genome Center (RCIGM-CGC) provides CAP/CLiA licensed clinical WGS testing and has a computational biobank of >2500 clinical genomes (and >4000 unaffected parental genomes for baseline). Genotyping CYP2D6 and other Pharmacogenomics genes is important for precision drug therapy because the enzyme it encodes metabolizes approximately 25% of drugs, and its activity varies considerably among individuals. Thus, patients on either end of the activity spectrum are at risk of subtherapeutic or supratherapeutic medication levels. Therefore, we aimed to assess the utility of the concurrent calling standard Pharmacogenomics genotyping (PGx) using WGS data for predicting opiate and other drug dosing for children in the Pediatric ICU. **Methods**: Subject genomes were interrogated by Illumina SBS chemistry (HiSeq, NovaSeq) to an average genome-wide coverage of >40x. Following the clinical bioinformatics WGS workflow (DRAGEN-based alignments to reference genome and variant calling, both small, indel, structural and copy number variants), read data was extracted for 34 PharmGKB level 1 gene loci (including but not limited to *CYP*-family, G6PD, RYR1, TPMT*) and *HLA* regions and processed with a combination of computational tools: Stargazer, PHLAT, OptiType (as per published protocols). The phenotype data for patients with computational prediction of poor/rapid metabolizer or unfavorable response was then collected and clinically correlated by targeted review of EPIC records for medical history and treatment and treatment responses. **Results**: An extensive dataset for precision and sensitivity evaluation was drawn from 1100 clinical genomes sequenced in 2020 at RCIGM-CGC. Computational analysis time was CYP3A4, and a wide variable spectrum for *CYP2D6* and the rest of pharmacogenes. Over 100 haplotypes (star alleles) have been detected for *CYP2D6*, some involving a gene conversion with nonfunctional paralog *CYP2D7*. Clinical correlation with neonatal ICU patients was minimal due to little overlap with age of onset, while higher in pediatric cohort (10+ yo). **Conclusion:** Accurate genotyping of pharmacogenes with WGS and subsequent allele calling with Stargazer, PHLAT, OptiType is feasible in clinical setting and will aid the implementation of precision drug therapy. In retrospective study, a small fraction of pediatric patients is identified, and clinical interaction is being followed up.

PrgmNr 3865 - Space-optimized HLA typing using Axiomâ¢ PGx research solutions

View session detail

Author Block: **C. Bruckner**; Thermo Fisher Scientific, Santa Clara, CA
Disclosure Block: **C. Bruckner:** Salary/Employment; Thermo Fisher Scientific.
The human leukocyte antigen (HLA) complex is the human version of the major histocompatibility complex (MHC). This complex includes genes responsible for immune function. Variations in these genes can affect adverse reactions to drugs, disease susceptibility, and immune response including transplant rejection. The highly polymorphic nature of this region and the prevalence of pseudogenes create challenges for traditional genotyping methods. Combining the use of direct genotyping with advanced imputation methods over the extended MHC region allows accurate HLA typing from SNP genotype data.
Applied Biosystemsâ¢ Axiomâ¢ microarrays are designed to measure approximately one thousand to nine hundred thousand markers as needed. Content prioritization is particularly important for space-constrained arrays. The Axiom Precision Medicine Diversity (PMD) Research Array genotypes over 850,000 markers, including around 8,000 for HLA typing. The Axiomâ¢ PharmacoFocusâ¢ array, designed for population-scale preemptive pharmacogenomics research, genotypes around 8,000 markers. HLA typing capability on Axiom PharmacoFocus Array was enabled by optimizing HLA module to around 1,500 markers. HLA marker selection was guided to enable efficiency while maintaining high HLA typing concordance by imputation, offering HLA typing across 11 MHC loci inclusive of: HLA-A*31:01, HLA-B*15:02, HLA-B*57:01, HLA-B*58:01.
A study used three Axiom microarrays with different density of HLA markers to genotype 1000 Genomes Project samples with diverse ancestries. The evaluated arrays are the PMD Array, PharmacoScanâ¢, and Axiom PharmacoFocus. HLA types for all 11 classical loci were imputed using the HLA*IMP:02 algorithm by Dilthey et al. (2013), as implemented in Axiom HLA Analysis software. High HLA typing concordance was maintained as the number of available markers was reduced by around 80% for the Axiom PharmacoFocus array. HLA types imputed by tested arrays also compared favorably to *in silico* HLA typing by Abi-Rached et al. (2018) of 1000 Genomes Project sequencing data.
References: [1] Dilthey AT et al. Multi-Population Classical HLA Type Imputation. PLoS Comput Biol. 2013; 9(2):e1002877. [2] Abi-Rached L et al. Immune diversity sheds light on missing variation in worldwide genetic diversity panels. PLoS One, 2018; 13(10):e0206512.

PrgmNr 3866 - The Study of *HLA-B*48:01* and *HLA-B*55:01* Frequency in Healthy Thai Population

View session detail

Author Block: **S. Chanuntranont**[1], A. Srisodsai[2], P. Satapornpong[3]; [1]Univ. of California, Riverside, Riverside, CA, [2]MedCoach Inst., Bangkok, Thailand, [3]The division of general pharmacy practice, Dept. of pharmaceutical care, Coll. of Pharmacy, Rangsit Univ., Pathum-Thani, Thailand
Disclosure Block: **S. Chanuntranont:** None.
Introduction: Penicillin, an antibiotic included in the beta-lactam (BL) class, is commonly used to treat bacterial infections in adults and children. It has been reported that approximately 10-20% of children who have received antibiotics in the beta-lactam class have developed a hypersensitivity reaction. In previous studies, it has been proven that *HLA-B*48:01* has an association with BL hypersensitivity in Thai children (OR = 37.4, 95%CI: 1.69-824.59; p = 0.016), and *HLA-B*55:01* has an association with penicillin hypersensitivity (OR = 1.41, 95%CI: 1.33-1.49; p = 2.04 $x10^{-31}$) in European ancestry. **Objective:** The aim of this study is to explore whether there is a distribution of *HLA-B*48:01* and *HLA-B*55:01* in healthy Thai individuals. **Methods:** 200 unrelated healthy Thai individuals were recruited in this study. *HLA-B* alleles using the Lifecodes HLA SSO typing kits (Immucor, West Avenue, Stamford, USA). **Results:** We found that the *HLA-B* alleles that show the most frequency in healthy Thai individuals were *HLA-B*46:01* (14.25%), *HLA-B*40:01* (7%), *HLA-B*58:01* (7%), *HLA-B*15:02* (6.75%), *HLA-B*13:01* (6.25%), *HLA-B*44:03* (4.75%), *HLA-B*51:01* (4.25%), *HLA-B*52:01* (4.00%), *HLA-B*35:05* (3.00%), and *HLA-B*40:06* (2.75%). Furthermore, there was not much distribution between *HLA-B*48:01* (0.25%) and *HLA-B*55:01*(0.25%) in Thais. Interestingly, the distribution of *HLA-B*48:01* was more prominent in Asian individuals: 2.88% of Japan, 3.56% of South Korea, 1.16% of Hong Kong, 1.00% of China, and 8% of Philippines. Whereas, *HLA-B*55:01* has 3.60% in Europeans and less than 1% in Asians. **Conclusion:** Thus, the distribution of pharmacogenetics markers could be used for screening among different ethnicities before initiation of penicillin treatment to avoid hypersensitivity reactions.

PrgmNr 3867 - A first exploration of the genetic architecture of linea nigra and its relationship with pigmentation

View session detail

Author Block: **S. Bivol**, J. Painter, P. Lind, C. Olsen, D. Whiteman, M. Law, S. Medland; QIMR Berghofer Med. Res. Inst., Brisbane, Australia
Disclosure Block: **S. Bivol:** None.
Introduction: Genome-wide association studies (GWAS) and follow-up meta-analyses have identified numerous genetic polymorphisms associated with melanocytic nevi (mole) count, and various pigmentation traits including hair, skin, and eye color. However, the genetic architecture of a common pregnancy hyperpigmentation - linea nigra - has not been explored before. Our aim was to conduct the first genetic analyses of this pregnancy complication and examine the genetic relationships with other pigmentation phenotypes. **Methods**: Phenotype data was collected via an online survey and DNA was extracted from saliva samples. The samples were genotyped on the Illumina Global Screening Array and imputed to the HRC reference panel. GWAS was conducted to explore genetic variations associated with linea nigra in a sample of 275 cases and 701 control Australian women of European descent. We computed a polygenic risk score (PRS) using data from a GWAS of mole count (N ~ 20,000, QSkin Sun & Health Study, QIMR Berghofer) to examine the relationship with linea nigra. **Results**: As expected given the low power, no genome-wide significant loci were found. However, several lead variants were nominally associated with linea nigra, including rs4734915 in *OXR*1 ($p = 4x10^{-6}$) and rs7634124 in *CLTN2* ($p = 7x10^{-6}$). Notably, *OXR1* and *CLTN2* genes have been associated with gestational age at birth and eye color, respectively, in previous GWAS analyses. The mole count PRS using a threshold of $1x10^{-6}$ significantly predicted linea nigra and accounted for 1.2% of the variance (P= 0.00057). **Conclusion:** A mole count PRS predicted linea nigra which suggests that linea nigra has shared genetic risk factors with melanocytic nevi. Although, no significantly associated loci were found, GWAS and PRS analyses will contribute to our understanding of the genetic architecture of linea nigra. Further work in larger samples is needed to identify causal loci for linea nigra and clarify its relationship with nevus count, pigmentation, and gestational age at birth.

PrgmNr 3868 - Clinical experience of an alpha thalassemia carrier screening assay with an increased detection rate due to novel variant calling

View session detail

Author Block: **G. Gould**, L. Collins, S. Pierson, R. Ben-Shachar, J. Wang, K. E. Erickson, K. Winters, K. Johansen Taber; Myriad Genetics, Inc., South San Francisco, CA
Disclosure Block: **G. Gould:** Salary/Employment; Myriad Genetics, Inc..
Background:
Carrier screening for alpha thalassemia, a prevalent genetic disease up to 200x more common than cystic fibrosis in certain ethnic populations, is recommended for all women who are pregnant or planning a pregnancy (1). Disease severity varies from mild anemia to in utero fetal demise. Most alpha thalassemia carrier screening assays detect only common copy number variants (CNVs) and the Constant Spring (CS) variant (2). To increase our detection rate, particularly across ethnicities to meet the needs of a diverse population, we expanded our assay to include novel variant calling of single nucleotide variants and insertions/deletions. Here we present the clinical experience of our improved assay.
Methods:
We collected data from 72,394 patients that received the Foresight Carrier Screen over a five-month period. We compared the percentage of patients that would have been identified as alpha thalassemia carriers before and after the addition of novel variant calling. Variants were classified according to case, functional and structural data, consistent with ACMG/AMP variant interpretation guidelines (3).
Results:
Detection rate improvement due to novel variants was highest in ethnicities that are frequently alpha thalassemia carriers. For example, ~8.5% of Middle Eastern patients were carriers of a CNV or CS while ~2% were carriers of a novel variant. Thus, novel variant calling increased carrier detection in the Middle Eastern population by >20%. Other ethnicities with the highest increases in detection rate include South Asian (4%), Southeast Asian (3%), East Asian (3%), Southern European (2%) and African/African American (1%).
In our cohort, we detected one novel variant - HBA2:c.95+2_95+6del5 - more often than CS, a well-characterized pathogenic variant. The c.95+2_95+6del5 variant is common in Mediterranean populations and leads to reduced Î±-globin expression. This variant produces a more severe phenotype than the corresponding deletion of the same gene. Clinically relevant phenotypes are seen in subjects with only two inactivated Î±-globin copies while deletional Hb H disease is only seen in subjects with 3 inactivated Î±-globin copies.
Conclusions:
Our results demonstrate novel variant calling for alpha thalassemia increases carrier detection rate. Despite the clinical importance of identifying novel variants, the inclusion of novel variant calling is not a routine part of all carrier screening and should be considered by healthcare providers.
References:
1.ACMG Committee on Genetics. Obstet. Gynecol. 129, 41-55 (2017).
2.Hogan et al. Clinical Chemistry 64:7 1063-1073 (2018).
3.Richards et al. Genet Med. 17(5):405-24 (2015).

PrgmNr 3869 - Developmental and temporal characteristics of clonal sperm mosaicism

View session detail

Author Block: **M. W. Breuss**[1,2,3], X. Yang[2,3], J. G. Gleeson[2,3]; [1]Univ. of Colorado, Sch. of Med., Aurora, CO, [2]Univ. of California, San Diego, La Jolla, CA, [3]Rady Children's Inst. for Genomic Med., San Diego, CA

Disclosure Block: **M.W. Breuss:** None.

Throughout development and aging, human cells accumulate mutations that result in genomic mosaicism and genetic diversity at the cellular level. Within the individual, in which mosaic variants arise, they have been connected to a range of diseases-including cancer or dementia. Those present in the gonads, however, can affect not only the individual, but also the offspring and subsequent generations. Indeed, we and others have demonstrated that pathogenic *de novo* mutations may be present as mosaic mutations in the germ cells of the father. Here we explore the patterns and temporal stability of clonal mosaic mutations in male gonads by sequencing ejaculated sperm and contrasting it with blood cells. We performed 300Ã whole-genome sequencing of both tissues from 17 healthy men, and we include assessment across multiple samples and age groups. We find that each male carries, on average, more than 30 clonal mosaic variants in their sperm, which are detected in serial sampling, and with the majority absent from sampled somal tissues. Across age groups, mosaicism detected in sperm remains stable, whereas blood mosaicism undergoes dynamic changes which are indicative of processes of clonal hematopoiesis. Moreover, two out of five individuals with advanced age harbored almost an order of magnitude more detectable mosaic variants in their blood than individuals of young age. The temporal stability and mutational signature of sperm mosaic variants suggest an origin during embryonic development, and a largely immutable stem cell niche during aging. Thus, clonal mosaicism in sperm is not correlated with the observed increase in mutational load in offspring of men with advanced paternal age. Nevertheless, this mosaicism contributes a transmissible, predicted pathogenic exonic variant for 1 in 15 men, representing a life-long threat of transmission for these individuals, and a significant burden on human population health. This knowledge allows for the development and implementation of prevention strategies for a subset of those disorders caused by *de novo* mutations; and it results in a framework that can counsel families with existing monogenetic disorders or prospective parents prior to conception.

PrgmNr 3870 - Exploring the use of whole genome sequencing for carrier screening

View session detail

Author Block: **X. Chen**[1], D. L. Perry[1], A. J. Coffey[2], N. J. Burns[1], R. J. Taft[1], M. A. Eberle[1]; [1]Illumina Inc., San Diego, CA, [2]Illumina United Kingdom, Cambridge, United Kingdom
Disclosure Block: **X. Chen:** Salary/Employment; Illumina Inc..
Advances in bioinformatic approaches provide an opportunity to conduct carrier screening via whole genome sequencing (WGS) by capturing clinically relevant conditions in a single assay and providing the ability to expand investigation across the genome. Here, we present an exploratory analysis using the 1000 Genomes Project dataset and retrospective analysis of clinical WGS cases with a diagnostic finding associated with an autosomal recessive (AR) or X-linked (XL) condition.
We developed a bioinformatics pipeline that filters variant calls genome-wide for ClinVar pathogenic (P) and likely pathogenic (LP) variants, as well as variants with severe consequences in ~2800 genes associated with known AR or XL diseases. Our pipeline also includes targeted variant calling for difficult genes, such as *FMR1* (fragile X syndrome), *SMN1* (spinal muscular atrophy), *HBA1*/*HBA2* (alpha thalassemia) and *CYP21A2* (21-Hydroxylase-Deficient Congenital Adrenal Hyperplasia). We applied the pipeline to 2504 individuals from the 1000 Genomes Project and found that ~80-90% of the individuals in each population carry at least one variant. Random pairing of male and female individuals revealed that 2.6% of the couples from the same ethnicity are at risk for an AR disorder, and 1.8% of the couples are at risk assuming pairing across all ethnicities.
Separately, we reviewed clinical WGS cases of patients with genetic diseases analyzed through the Illumina Clinical Services Laboratory and the iHope Program. Of 826 cases, 96 (11.6%) had a diagnostic variant (P or LP) in a gene associated with an AR or XL disorder. Of these, 44 were due to compound heterozygous variants and 38 were due to a homozygous variant associated with an AR disorder, 10 were due to a hemizygous variant associated with an XL disorder, and 4 were due to a heterozygous variant associated with an XL disorder (females). Review of these variants in the context of our carrier screening pipeline described above indicates that 75.3% of individual parents would have been identified as carriers and 97.9% of couples could be identified with more lenient filtering criteria. Preliminary retrospective analysis of 51 of these variants indicates that 48 (94%) would reach an LP or P classification in the parents alone and 3 variants (homozygous and classified as LP in the proband) would be classified as uncertain significance (VUS) due to inability to apply ACMG criteria relating to phenotype and segregation.
WGS-based carrier screening with the integration of targeted callers has the potential to provide a comprehensive carrier screening analysis in a single assay. Future work includes refinement of an interpretation and reporting workflow.

PrgmNr 3871 - Genematching candidates from the Genomic Autopsy Study to elevate variant classification for prenatal clinical use

View session detail

Author Block: **T. Ha**[1,2], P. Arts[1], A. B. Bryne[3], J. Feng[1,2], D. Lawrence[1], M. Babic[1], L. Pais[3], Genomic Autopsy Study Research Network, S. L. King Smith[1], M. R. Jackson[1,4], A. W. Schreiber[2,5], A. O'Donnell-Luria[3], C. P. Barnett[6,7], H. S. Scott[1,2,4,7]; [1]Dept. of Genetics and Molecular Pathology, Ctr. for Cancer Biology, an alliance between SA Pathology and the Univ. of South Australia, Adelaide, Australia, [2]ACRF Genomics Facility, Ctr. for Cancer Biology, An alliance between SA Pathology and the Univ. of South Australia, Adelaide, Australia, [3]Ctr. for Mendelian Genomics, Broad Inst. of MIT and Harvard, Cambridge, MA, [4]Australian Genomics, Melbourne, Australia, [5]Sch. of Biological Sci., Univ. of Adelaide, Adelaide, Australia, [6]Paediatric and Reproductive Genetics Unit, South Australian Clinical Genetics Service, Womenâ□□s and Childrenâ□□s Hosp., North Adelaide, Australia, [7]Sch. of Med., Univ. of Adelaide, Adelaide, Australia

Disclosure Block: **T. Ha:** None.

Background: Trio analysis from the Genomic Autopsy Study revealed that a molecular diagnosis can be ascertained in half of clinically unresolved cases of early miscarriage or perinatal death (n=89/170). However, only half (n=41/89) of these putative disease genes can be used prenatally for assessing recurrence risk, despite compelling evidence from available animal models, gene constraints and expression data. The ACMG guidelines are heavily weighted towards existing genotype to phenotype delineations in the post-natal setting, which can confound the interpretation of prenatal cases, with little to no hindsight on severe or lethal in utero presentations[1].

Aim: Assess the proportion of cases with variants and/or genes of uncertain significance, which can be reclassified based on identifying additional kindreds from genematching.

Methods: Between 2018 to 2021, we submitted 35 genes to the genotype matching platform, MatchMaker Exchange, to seek additional, unrelated kindreds with similar clinical phenotype and/or initiate collaborations with expert curators of genes or disease subgroups for each case[2].

Results: We received matches for 33/34 genes submitted, with an average 7 participants responding per gene and contact initiated between 1-25 months after submission. The genematching identified additional kindreds with clinical overlap (10/33) and/or initiated functional validation (5/33) and diagnosis for publications (3/33).

Conclusion: Phenotype matching is critical for clinical interpretation, which can be accelerated through the use of genotype sharing platforms like MatchMaker exchange. Additional kindreds can be identified in a fifth of cases with variants or genes of uncertain significance.

References:

1.Monaghan, K. G., Leach, N. T., Pekarek, D., Prasad, P. & Rose, N. C. The use of fetal exome sequencing in prenatal diagnosis: a points to consider document of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med*. **22**, (2020).

2.Azzariti, D. R. & Hamosh, A. Genomic Data Sharing for Novel Mendelian Disease Gene Discovery: The Matchmaker Exchange. *Annu. Rev. Genomics Hum. Genet*. **21**, (2020).

PrgmNr 3872 - Leveraging variant classification evidence from diagnostic genetic testing to improve the rate of positive results in reproductive carrier screening

View session detail

Author Block: **J. Wilkinson**, S. Garcia, A. Mitchell, S. Poll, S. Aradhya; Invitae, San Francisco, CA
Disclosure Block: **J. Wilkinson:** Salary/Employment; Invitae.
Background: With the increasing use of expanded carrier screening (ECS) for hundreds of genes, accurate and definitive classification of some variants becomes a major challenge as available evidence to interpret clinical significance may not be sufficient. In such cases, additional evidence may be obtained through diagnostic genetic testing of individuals who are clinically affected and have those same variants. Advances in variant classification that incorporate evidence from a functional modeling platform (FMP) for missense variants, and from applying clinical pathognomonic criteria for characteristic diseases, have enhanced diagnostic testing. This study aimed to quantify this increase in positive carrier screening results as a result of leveraging diagnostic testing data and improved variant classification.
Methods: Reproductive carrier screening was performed for 134,841 individuals for up to 301 genes. We collected positive carrier variants which used internal case reports that met pathognomonic criteria or evidence from FMP. Data was then matched to ClinVar to identify instances in which this internal data constituted additional evidence for variant classification. Only substitutions and indels were considered for this analysis to ensure accurate matching.
Results: A total of 59,494 individuals were carriers for one or more pathogenic or likely pathogenic substitutions or indels. There were 3,071 unique variants, in 18,259 (28%) individuals, for which the clinical classification depended in part on evidence from pathognomonic criteria or FMP. While ~25% of these were truncating variants and therefore easier to classify as pathogenic, 924 unique variants in 2,519 individuals (~4% of cases) were uniquely classified using novel information derived from diagnostic testing.
Discussion: As larger numbers of individuals pursue expanded carrier screening, many more novel variants will require accurate clinical classification. By leveraging data from diagnostic testing performed for the same genes in affected individuals, and by appropriately applying pathognomonic criteria and FMP evidence to variant interpretation, informative results for reproductive risk assessment can be provided to a significantly higher number of individuals undergoing ECS.

PrgmNr 3873 - Male microchimerism in females: A quantitative study of twin pedigrees to investigate mechanisms

View session detail

Author Block: **B. Johnson**[1], H. E. Peters[2], C. B. Lambalk[2], C. V. Dolan[3], G. Willemsen[3], L. Ligthart[3], V. Mijatovic[2], J-J. Hottenga[3], E. A. Ehli[1], D. I. Boomsma[3]; [1]Avera Inst. for Human Genetics, Sioux Falls, SD, [2]Amsterdam UMC, Amsterdam, Netherlands, [3]Vrije Univ. Amsterdam, Amsterdam, Netherlands
Disclosure Block: **B. Johnson:** None.
Background: The occurrence of natural chimerism has been most commonly identified as microchimerism (Hypothesis: We hypothesized that characteristics of an individualâ□□s pedigree, such as having a male co-twin, older brothers, or sons, would increase an individualâ□□s risk of presenting with male microchimerism.
Study design, materials, methods: To explore this complex phenomenon we have employed a twin pedigree study to investigate the prevalence and patterns of male microchimerism within families. This study includes 446 adult female participants recruited by the Netherlands Twin Register (NTR), including female twins, mothers and sisters of monozygotic, same-sex dizygotic, and opposite-sex dizygotic twin pairs. Biobanked DNA isolated from peripheral blood samples was tested via a high sensitivity quantitative PCR assay to measure male specific Y chromosome gene *DYS14* relative to total measure of a beta-globin gene *HBB*.
Results: This assay produced detectable measure of male microchimerism in 26.9% of participants with similar median concentrations of male genome among types of twins, their sisters and their mothers (P = 0.28). Pedigree analysis found the prevalence of male microchimerism is not explained by having male offspring (OR = 0.90: SE 0.19, P = 0.63), an older brother (OR = 1.46: SE 0.32, P = 0.09) or a male co-twin (OR = 1.23: SE 0.40, P = 0.61). Correlation among monozygotic twin pairs (0.27; SE 0.37) and first-degree relatives (0.091; SE 0.092) is statistically similar (P = 0.66). There is a significant positive relationship between the age at time of biobanking and detection of male microchimerism (P = 0.02; Nagelkerke R2 = 0.017), however quantitative outliers (>80 male genome equivalents) among positive participants present with a lower age at time of biobanking (P = 0.034).
Implications: While this study does not reveal a specific pedigree characteristic that promotes male microchimerism prevalence, this may support time, molecular traits, and environmental exposures as contributors to the general prevalence of persistent microchimerism. This study further necessitates study of molecular underpinnings of natural microchimerism to understand implications on immunology, autoimmune diseases, and overall womenâ□□s health.

PrgmNr 3874 - Whole-exome sequencing confirms GDF15 is the greatest genetic risk factor for Hyperemesis Gravidarum and identifies rare and novel variants in appetite genes GDF15 and GFRAL

View session detail

Author Block: **M. Fejzo**[1], K. MacGibbon[2], P. Mullin[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]Hyperemesis Ed. and Res. Fndn., Clackamas, OR
Disclosure Block: **M. Fejzo:** Consultant/Consulting Fees/Other Remuneration; Materna Biosciences, Inc.
Many people experience nausea and vomiting of pregnancy and 18% require medication. The most severe form, Hyperemesis Gravidarum (HG), occurs in approximately 2% of pregnancies and is associated with weight-loss and undernutrition. HG can cause serious maternal morbidity and lead to suicidal ideation and PTSD. It is associated with adverse outcomes including preterm birth, neurodevelopmental delay, and autism spectrum disorder. It is highly heritable. Previously, our GWAS showed the placenta, appetite, and cachexia gene *GDF15* is the greatest genetic risk factor for HG. *GFRAL*, encoding the receptor for GDF15, is also associated. Herein we report results from whole-exome sequencing (WES) of 888 cases with HG, 661 unaffected controls, and 31 families. WES confirmed the paradigm-changing association between *GDF15* and HG- the *H6D* variant, previously associated with body weight and abdominal fat in a cachexia mouse model, was the only variant to reach genome-wide significance ($p=9.98x10^{-11}$). To identify rare variants in *GDF15* and *GFRAL*, variants were filtered and annotated to include rare (global frequency 0.05), non-synonymous variants unique to cases. Ancestry/relationships were predicted. Non-segregating variants were removed. Rare variants in *GDF15* and *GFRAL* were identified in 23 and 17 cases, respectively. Notably, in *GDF15*, 11 cases of European ancestry had variant C211G including a mother/daughter. In addition, 2 novel missense variants predicted to be damaging in patients of European ancestry (R193C, mature peptide cleavage site; C203S, mature peptide) and 1 novel stopgain variant (S174*) in a patient of African ancestry were identified. Six cases of 5 ancestries had upstream variant rs77109188. In *GFRAL*, 3 damaging missense variants were identified. M370T (transmembrane domain) was found in a case of European/American and a case of African ancestry, and variants in the GDF15 binding domain (E203A, Q184R) were found in cases of European ancestry.
This study confirms *GDF15* is the greatest genetic risk factor for HG and identifies rare variants in *GDF15* and *GFRAL*, providing further evidence the hormone pathway plays a role in disease etiology. Drugs targeting this pathway are currently under intense investigation to potentially treat obesity and conversely, cancer cachexia. Further work to explore the function of these variants will help elucidate molecular mechanisms of HG and provide clues to developing treatments for disorders of appetite, nausea, and vomiting.

PrgmNr 3875 - A variational Bayesian approach to characterize latent genetic components using GWAS summary statistics

View session detail

Author Block: **Z. Zhang**[1], J. Jung[1], N. Suboc[1], S. Gazal[1], N. Mancuso[2,3,4,1]; [1]Ctr. for Genetic Epidemiology, Dept. of Preventive Med., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [2]Univ. of Southern California, Los Angeles, CA, [3]Dept. of Quantitative and Computational Biology, Univ. of Southern California, Los Angeles, CA, [4]Norris Comprehensive Cancer Ctr., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA
Disclosure Block: **Z. Zhang:** None.
Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with multiple phenotypes, providing opportunities to gain insight into their biological function. Recent work aiming to characterize these shared latent genetic factors using truncated singular-value decomposition (tSVD) of GWAS summary data has been applied to numerous traits, however its model-free approach leaves unclear 1) what genetic parameters are modelled and 2) how uncertainty in observed summary data influences estimates.
To address these limitations, we derive from first principles a factor analysis model to identify latent genetic factors shared across phenotypes using GWAS summary data. Our approach accounts for uncertainty in effect-size estimates and uses an automatic relevance determination (ARD) prior to prune uninformative factors. In simulations, we found our approach identifies the true number of latent factors and outperforms tSVD in reconstruction error of unobserved allelic effect sizes. We observed that by directly modelling standard errors, unlike tSVD, our approach remains unbiased when there is differential power across GWAS studies.
We applied our approach to GWAS summary data from 158 metabolites in the Finnish Metabolic Syndrome in Men cohort (N=6263). To explore the functional roles of inferred latent genetic factors, we performed a phenotype enrichment analysis using Phenome-wide Association Studies (PheWAS) catalogues weighted by factor loadings. Focusing on the top 10 factors, we found that loadings from our method identified significantly more PheWAS phenotypes (182) in contrast to tSVD (99; P=9.97E-8). For example, the fourth factor from our method is enriched for SNPs associated with cervical human papillomavirus (HPV), uveitis, disturbance in tooth eruption, and respiratory diseases (FDR Padj=0.03), which is consistent with shared genetic components related to inflammation.
We found the number of identified phenotypes decreased per latent factor when ordered by inferred ARD parameters (P=7.35E-4), with little evidence when using tSVD orderings (P=0.07). This suggests that latent genetic factors identified by our approach better reflect biological informativeness compared with tSVD factors.
Taken together, our results suggest that our approach prioritizes biologically meaningful genetic factors underlying observed GWAS signals.

PrgmNr 3876 - Biobank-scale estimation of the proportion of trait variance explained by gene-environment interactions

View session detail

Author Block: **A. Pazokitoroudi**[1], A. Dahl[2], N. A. Zaitlen[3], S. Rosset[4], S. Sankararaman[3]; [1]Computer Sci., UCLA, Los Angeles, CA, [2]Univ. of Chicago, Chicago, IL, [3]UCLA, Los Angeles, CA, [4]Tel Aviv, Israel
Disclosure Block: **A. Pazokitoroudi:** None.

Studies of gene-environment interactions (GxE) aim to demonstrate how environmental factors and genetic markers jointly explain complex traits. Understanding the role of GxE has the potential to provide insights into mechanisms and pathways underlying disease risk, explain sources of heritability, and improve the accuracy of genetic risk prediction. The availability of large biobanks that measure both genetics and diverse phenotypes offers the promise of obtaining novel insights into GxE. However, current methods for polygenic GxE inference, typically based on maximizing the likelihood in a linear mixed model (LMM), do not scale to the biobank setting, where hundreds of thousands of individuals are measured across hundreds of traits with accompanying genotypes over millions of SNPs.

In this work, we propose a scalable algorithm, GENIE, to jointly estimate the proportion of trait variance that can be explained by genetics, environment, and gene-by-environment interactions, which can improve GxE estimates and eliminate false positives when the assumptions of existing methods are not satisfied. Our algorithm is a randomized method-of-moments estimator that has a runtime linear in each of the number of environmental variables and the number of SNPs while scaling linear in the number of individuals. We evaluate the accuracy and scalability of GENIE for estimating the variance components under different settings. In small-scale simulations, where GENIE can be compared to other computationally intensive estimators, we show that GENIE obtains estimates of GxE that are of accuracy comparable to maximum-likelihood estimates. More importantly, GENIE is highly scalable and can estimate GxE on 300K individuals, 500K SNPs and 10 environments in a few hours, a setting where established tools fail.

We applied GENIE to estimate the proportion of phenotypic variance explained by gene-by-statin interaction for 50 quantitative phenotypes (5 anthropometric, 45 blood biomarkers) with genotypes from 300,000 unrelated white British individuals. Across the fifty traits, we observed statistically significant evidence for gene-by-statin interaction (p

PrgmNr 3877 - Capturing fine-scale structure of single-cell expression and chromatin accessibility using a topic model

View session detail

Author Block: **P. S. Carbonetto**, K. Luo, A. Sarkar, A. Hung, S. Pott, M. Stephens; Univ. of Chicago, Chicago, IL
Disclosure Block: **P.S. Carbonetto:** None.
Grade of membership models have been widely used to model admixture in population genetics and to identify topics in document clustering applications. Recent papers have highlighted the potential for grade of membership models to identify structure from single-cell RNA-seq and ATAC-seq data, and in particular structure that is not well captured by clustering. However, important methodological and practical barriers remain, including the high computational cost of model fitting, and the question of how to interpret the structure identified by these models. To address these challenges, we developed a new approach for efficiently fitting grade of membership models to large data sets, and we developed extensions of differential expression and enrichment analysis that allow for partial membership of cells to topics. Applying this approach to published single-cell RNA-seq and ATAC-seq data sets, we identified both discrete cell types and continuously varying or transitional states that would not be easily captured by conventional methods. For example, in a population of mouse epithelial cells we identified rare transitional *Krt13*+ (â□□hillockâ□□) cells. These hillock cells were originally discovered via an ad hoc two-stage analysis using diffusion maps, whereas here they emerged as a single topic in the grade of membership model. Our analysis of single-cell chromatin accessibility profiles of kidney cells suggests that some specialized kidney cells exist as discrete cell types, whereas others vary along a spectrum, most strikingly among distal tubule and collecting duct cells. In a population of hematopoietic cells, we identified continuous differentiation processes among and within hematopoietic progenitor cell types that would be missed by clustering. In summary, we have developed a unified topic-model-based analysis pipeline for single-cell RNA-seq and ATAC-seq data that is simple, scalable, and can flexibly capture different types of structure in cell populations. Our methods are implemented in the R package fastTopics.

PrgmNr 3878 - Combining SNP-to-gene linking strategies to pinpoint disease genes and assess disease omnigenicity

View session detail

Author Block: **S. Gazal**[1], O. Weissbrod[2], F. Hormozdiari[3], K. Dey[4], J. Nasser[5], K. Jagadeesh[5], D. Weiner[5], H. Shi[4], C. Fulco[5], L. O'Connor[5], B. Pasaniuc[6], J. Engreitz[7], A. L. Price[4]; [1]USC, Los Angeles, CA, [2]Modi'in-Macabim-Reut, Israel, [3]Google Hlth., Cambridge, MA, [4]Harvard T.H. Chan Sch. of Publ. Hlth., Boston, MA, [5]Broad Inst., Cambridge, MA, [6]Geffen Sch. of Med. at UCLA, Los Angeles, CA, [7]Stanford Univ., Stanford, CA

Disclosure Block: **S. Gazal:** None.

Although genome-wide association studies (GWAS) have identified thousands of disease-associated common SNPs, these SNPs generally do not implicate the underlying mechanisms and target genes, as most disease SNPs are regulatory with unknown function. Many SNP-to-gene (S2G) linking strategies have been proposed to infer the function of regulatory SNPs and link them to the genes that they regulate, but it is unclear which of these strategies should be prioritized in the context of human disease.

We developed a framework for evaluating and combining different S2G strategies to optimize their informativeness for human disease, using polygenic analyses of disease heritability. We define an S2G strategyâ□□s *precision* and *recall* for identifying disease genes by evaluating the heritability enrichment of SNPs linked to a critical gene set (e.g. genes with highly constrained exons and promoters), relying on the hypothesis that a precise S2G strategy should maximize the heritability linked to the critical gene set compared to the heritability linked to all genes; recall is computed using the heritability linked to all genes. We applied our framework to GWAS summary statistics from 63 diseases and complex traits (average N=314K), evaluating 50 S2G strategies, and validated our precision and recall metrics using gold standard sets of causal SNP-gene pairs. Our optimal combined S2G strategy included 7 S2G strategies (Exon, Promoter, 2 fine-mapped cis-eQTL strategies, EpiMap enhancer-gene linking, Activity-By-Contact (ABC), and Cicero), and achieved a precision of 0.75 and a recall of 0.33, more than doubling the precision and/or recall of any individual strategy.

We describe two applications of our combined S2G strategy. First, we applied our optimal combined S2G strategy to fine-mapping results for 49 UK Biobank traits to predict 7,111 causal SNP-gene-disease triplets (with S2G-derived functional interpretation) with high confidence. Second, we applied our optimal combined S2G strategy to genome-wide fine-mapping results for these traits (not restricted to GWAS loci) to rank genes by the heritability linked to each gene, providing an empirical assessment of disease *omnigenicity* (Boyle et al. 2017 *Cell*); averaging across traits, we determined that the top 200 (1%) of ranked genes explained roughly half of the heritability linked to all genes.

Our results highlight the benefits of combining different S2G strategies to pinpoint disease genes. We anticipate that precision and recall will increase further under our framework as improved functional assays lead to improved S2G strategies.

PrgmNr 3879 - Development of a Polygenic Risk Score to Predict Severe Outcome of a COVID19 Infection

View session detail

Author Block: **D. E. Condon**, P. Cherukuri, S. Gu, L. Carmichael, C. Hajek; Sanford Hlth., Sioux Falls, SD
Disclosure Block: **D.E. Condon:** None.
COVID19 is a contagious respiratory viral infection. Patient outcome varies from an asymptomatic viral infection to a severe pulmonary infection, possibly requiring respiratory support and even to death. Our study leverages machine learning using machine learning to predict which patients that test positive for COVID-19 are likely to require enhanced care, which we define as hospitalization and/or use of respiratory support. Sanford Healthâ□□s Enterprise Data Analytics created an index score within [0,1] that directly correlates with a patientâ□□s likelihood to require hospitalization. The index score is based on known risk factors for an adverse event during COVID-19 infection, such as age, excessive BMI, male sex, known comorbidities, etc. We aim to improve on the index score by including genetic data as measured by Illuminaâ□□s Global Screening Array (GSA) version 1 to construct a polygenic risk score (PRS). Clinical data was extracted from the medical record and de-identified for approximately 11,000 patients that have tested positive for COVID-19 within the Sanford Health system. We use a combination of known alleles that can be directly measured by the GSA and proxy assays for alleles that are not present on the GSA to create different PRSs: 1) a conventional PRS which uses a sum of risk allele counts, 2) a conventional weighted PRS, and 3) a non-standard PRS that attempts to account for Mendelian inheritance, i.e. where 0 and 1 risk alleles, or 1 and 2, could have the same effect, 4) XGBoostClassifier machine learning based prediction of outcome likelihood. The results are compared, and proposed as a tool to supplement physician decision-making.

PrgmNr 3880 - Estimating heritability explained by local ancestry and evaluating stratification bias in admixture mapping from summary statistics

View session detail

Author Block: **T. Chan**[1], C. W. Chiang[1,2], N. Mancuso[1,2,3]; [1]Ctr. for Genetic Epidemiology, Dept. of Preventive Med., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [2]Dept. of Quantitative and Computational Biology, Univ. of Southern California, Los Angeles, CA, [3]Norris Comprehensive Cancer Ctr., Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA
Disclosure Block: **T. Chan:** None.
Admixture mapping aims to identify genomic regions associated with a phenotype through the ancestry components of an admixed population. It leverages the fact that local ancestry can tag underlying causal genetic variants, both common and rare, that differ in allele frequencies in the ancestral populations. The proportion of heritability explained by local ancestry, or h_A^2, can provide insights into the latent genetic architecture by proxy for narrow-sense heritability when underlying causal variants are poorly genotyped or imputed in GWAS. Indeed, recent work has demonstrated that narrow-sense heritability can be inferred directly as a function of h_A^2, this approach produces an estimate from unrelated individuals while capturing the heritability due to uncommon variants that is missing from genotyped SNP heritability. However, we identify two primary limitations with this approach. First, estimating h_A^2 from individual-level data is computationally limiting due to likelihood evaluation in linear mixed models and relatedness estimation. Second, estimates of h_A^2 may be susceptible to biases due to stratification in ancestral populations. To address these limitations, here we propose a computationally efficient approach to infer h_A^2 from admixture mapping summary statistics that is robust to ancestral population stratification. Our method models the genome-wide covariance across local ancestry markers using a computationally efficient truncated SVD approach while maintaining unbiased estimation of model parameters. Using extensive population genetic simulations, we demonstrate that ancestral stratification inflates estimated effect sizes at local ancestries and biases estimates of h_A^2 in individual-level data. We show that our approach provides approximately unbiased estimates of h_A^2 (slope=1.00, 95%CI: 0.90, 1.10) across a variety of trait architectures and ancestral demographies. As an illustration of downstream applications, we leverage our model to perform a genetic-architecture-aware sampling procedure to infer the genome-wide significance threshold for admixture mapping. We demonstrate that our improved sampling procedure adequately controls the FWER under a variety of genetic architectures and demographies, unlike previous approaches. For example, when 16% of phenotypic variance is explained by stratification, our approach controls FWER at 5% compared with 13% when using previous sampling procedures. Overall, our approach utilizes in-sample summary statistics and admixture history to estimate h_A^2 that is robust to stratification bias.

PrgmNr 3881 - Explainable and extendable machine learning models for identifying prognostic radiogenomic biomarkers from breast cancer multimodal imaging and genomic data

View session detail

Author Block: **Q. Liu**, P. Hu; Univ. of Manitoba, Winnipeg, MB, Canada
Disclosure Block: **Q. Liu:** None.
Background: Radiogenomics is a field where medical images and genomic profiles are jointly analyzed to answer critical clinical questions. We proposed a novel framework to identify prognostic radiogenomic biomarkers from multi-modal breast cancer (BC) magnetic resonance imaging (MRI) and multi-omics data, which may serve as a substitute for genetic testing.
Methods: Bayesian tensor factorization (BTF) was used to extract the integrated multi-omics features from gene expression, DNA methylation, and copy number variation data of 762 BC patients. The potential biological functions of these BTF multi-omics features were explored using Gene set enrichment analysis (GSEA). A deep learning (DL)-based imaging segmentation model was built to extract multi-modal MRI radiomic features for 61 of the BC patients with MRI data. Two explainable tools (Gradient and Gradient*Input) were embedded into the DL model structure to explore biological implications of the radiomic features. Predictive least absolute shrinkage and selection operator (LASSO) models were trained to translate the radiomic features from the BTF multi-omics features for the BC patients without MRI data. Survival analyses were then performed to estimate the prognostic significance of each radiomic feature. Statistical mediation analyses were performed to further explore the underlying biological mechanisms of the identified biomarkers. Traditional semi-auto radiomic features and previously established single-omics features (e.g., BC risk gene expressions, pathway activity scores, and gene signature scores calculated from the gene expression profile) were used as baselines for comparison.
Results: Saliency maps of the multi-modal MRI radiomic features showed the excellent explainability of the built DL models. Three DL-based multi-modal MRI radiogenomic biomarkers were successfully identified. which were confirmed to have significant differences in overall survival (log-rank test, Bonferroni corrected *P* valueAPITD1, *HNF4*) and several metabolism related pathways (Purine metabolism pathway and Tryptophan metabolism pathway), which has a significant mediation effect on the relationship between one specific BTF multi-omics feature, representing the function of natural killer cells based on the GSEA analysis, and the BC survival time (adjusted *P* value**Conclusion:** The results may promote MRI as a non-invasive examination for BC prognosis and multi-level molecular status, and ultimately increase precision in BC prognosis and improve patient care.

PrgmNr 3883 - FDR confidence interval selection and adjustment approach for large-scale hypothesis testing

View session detail

Author Block: **J. Millstein**[1], F. Battaglin[1], H. Arai[1], W. Zhang[1], P. Jaychandran[1], S. Soni[1], A. R. Parikh[2], C. Mancao[1], H-J. Lenz[1]; [1]Keck Sch. of Med. of USC, Los Angeles, CA, [2]Massachusetts Gen. Hosp., Boston, MA

Disclosure Block: **J. Millstein:** None.

Momentum has been gaining among statisticians in recent years to go beyond the dichotomous and somewhat arbitrary dictates of the *p*

PrgmNr 3884 - iCURL: Refining Estimated Identical-by-Descent Segments Via Community Detection

View session detail

Author Block: **R. Shemirani**[1], C. R. Gignoux[2], N. A. Zaitlen[3,4], C. L. Avery[5], E. Kenny[6], G. M. Belbin[7], J. Ambite[8]; [1]Icahn Sch. of Med. at Mount Sinai, Marina del Rey, NY, [2]Univ of Colorado Denver, Anschutz Med. Campus, Aurora, CO, [3]UCSF, San Francisco, CA, [4]Univ. of California Los Angeles, Los Angeles, CA, [5]Univ. of North Carolina at Chapel Hill, Chapel Hill, NC, [6]Icahn Sch. of Med. at Mt Sinai, New York, NY, [7]Icahn Sch. of Med. at Mount Sinai, New York, NY, [8]Univ. of Southern California, MARINA DEL REY, CA
Disclosure Block: **R. Shemirani:** None.

Detection of segments of the genome shared among individuals identical-by-descent (IBD) is an essential step in many genomic analysis pipelines. New IBD estimation techniques are able to process hundreds of thousands of samples in large biobanks using phased genotype data and hashing heuristics. However, the resolution of IBD estimation in these approaches is limited by factors such as the density of the genotype array and phasing errors which may increase the rate of false-positive and false-negative results. Clustering algorithms have been previously shown to be able to address these challenges. Applying clustering algorithms to local IBD segments works as follows: first, each chromosome is divided into short windows. Secondly, for each window, a local IBD graph is created where nodes represent sample haplotypes and edges represent IBD sharing among individuals in that window. Ideally, in the absence of false-positive and false-negative edges, these graphs should solely be composed of cliques, that is, groups of individuals who share the same IBD-haplotype homologously from a recent common ancestor. Clustering algorithms help recover the cliques at each window. The resulting set of edges provides a fine-grained refinement of the IBD estimation results, where false-negative edges are interpolated and false-positive edges are removed. Previous clustering algorithms used to analyze local IBD graphs are not scalable and depend on *a priori* knowledge about IBD segments. Here, we introduce a scalable pipeline for local IBD clustering, iCURL (Identity-by-descent Clustering of Related Loci). Utilizing high-performance computing clusters, iCURL analyzes the entirety of UK Biobank IBD data in an hour (N=487,330 individuals, n=14,402,404,038 pairwise IBD-haplotypes). Our algorithm could also estimate the number of false-positive and false-negative IBD segments in each window. Applying our method to the UK Biobank, we found a total of 975,646,819 local IBD clusters. We evaluated the results of iCURL using whole-exome sequence data. Local IBD clusters generated with iCURL recovered 23% of rare variants with minor allele counts of 2 or 3 in the whole-exome sequence data. This recovery rate decreased as the minor allele frequency increased, replicating the results of other high-resolution IBD estimation approaches. These clustered IBD segments can be utilized for a variety of downstream analyses, including IBD mapping applications, to search for causal rare variants not captured in array data. We conducted IBD mapping analysis on a number of phenotypes successfully to further evaluate the results of iCURL and report on novel associations with height and lipids.

PrgmNr 3885 - Improving Polygenic Risk Score Transferability by Leveraging GWAS Signals from Diverse Populations

View session detail

Author Block: **J. Shi**[1], W. Wang[2], S. S. Shringarpure[3], 23andMe, Inc. Research Team, A. Auton[4]; [1]23andMe Inc., Sunnyvale, CA, [2]Santa Clara, CA, [3]23andMe, Sunnyvale, CA, [4]23andMe, Mountain View, CA

Disclosure Block: **J. Shi:** Salary/Employment; 23andMe Inc..

Polygenic risk scores (PRS) from large scale genome-wide association studies (GWAS) can provide useful information for personalized risk stratification and disease risk assessment. However, to date, GWAS are predominantly based on European populations, trailed by Asian populations, while all other ancestry groups comprise less than 5%. PRS based on European training data suffer reduced accuracy in non-European target populations, exacerbating health disparities and limited generalizability across other ancestral populations. Here, we propose MultiPRS, a method that improves trans-ethnic polygenic prediction by borrowing strength from GWAS signals in multiple populations. To construct MultiPRS for a phenotype in a given (non-European) population, we first applied the Pruning and Thresholding (P+T) method (with various P-value and LD-clumping thresholds using population-specific LD panels) to perform variant selection on GWAS summary statistics in each of the European, Latino, African American, East Asian and South Asian ancestries. We then constructed ancestry-specific P+T PRS as the weighted sum of the target populationâ□□s imputed dosages, where the weights were GWAS effect size estimates from the P+T selected variants in each of the five populations. Lastly, we applied logistic regression with elastic-net regularization on the ensemble of P+T PRS across populations and parameter settings to obtain the final MultiPRS. We assessed the MultiPRS prediction performance in the African American population on asthma, chronic kidney disease, gout, T2D and uterine fibroids. Compared with PRS that were purely based on the African American GWAS summary statistics, our approach led to approximately a 6.87% AUC relative increase on average, ranging from 3.15% to 11.55%. For example, the AUC of T2D PRS increased from 0.603 (95% CI: 0.591 - 0.615) to 0.652 (95% CI: 0.640 - 0.664) by transferring GWAS signals from the other populations. MultiPRS also yielded similar AUC compared to PRS constructed from the trans-ethnic meta-analysis summary statistics (e.g., for T2D, MultiPRS AUC = 0.652 v.s meta-analysis PRS AUC = 0.650), but to further boost prediction power, one can easily add additional PRS (from genetically correlated phenotypes with stronger GWAS signals) in the elastic-net ensemble step of MultiPRS, which is not feasible if using the trans-ethnic meta-analysis summary statistics.

PrgmNr 3887 - Repeat Expansions and Somatic Instability Observation using Optical Mapping

View session detail

Author Block: **A. Hastie**[1], J. Lee[2], A. Pang[1], D. Zhang[1], H. Houlden[3], A. Chaubey[4]; [1]Bionano Genomics, San Diego, CA, [2]BioNano Genomics, San Diego, CA, [3]Inst Neurology, London, United Kingdom, [4]Greenwood Gen Ctr, Greenwood, SC
Disclosure Block: **A. Hastie:** Salary/Employment; Bionano Genomics.
Tandem repeats are expansions of a sequence forming a repetitive pattern in the genome. These regions are often unstable and can be associated with genetic disorders, with the size of expansions correlating with the severity and age of onset. Therefore, being able to accurately detect the total length of expansion and any somatic expansions is important. Here we show that using intact high molecular weight optical genome mapping (OGM) molecules, the length of stable and mosaic repeat expansions across the genome can be revealed efficiently. PCR based techniques are sufficient for sizing of low range expansions, but expansions over a few hundred base pairs are challenging due to polymerase slippage and stuttering. Southern blot is currently the gold standard for longer repeat length detection, but it is locus-specific, laborious and expensive. Using OGM, we can select within the reference genome the two labels flanking the repeats of interest, measure the interval lengths in the aligned molecules, and estimate the mean and variance of the repeat sizes. For example, a mosaic repeat expansion with a wide distribution of repeat lengths from 7 kbp to 17 kbp is observed in an amyotrophic lateral sclerosis (ALS) sample. Using the label distances in the optical mapping molecules, a histogram of repeat sizes and a Gaussian mixture model can be used to identify the zygosity of the repeat expansion region. Moreover, with the standard deviations of the clusters, a mosaic expansion of various repeat sizes is observed. This method has also been applied to other repeat expansion disorders such as myotonic dystrophy (DM), Spinocerebellar ataxias (SCAs) and Fragile X syndrome (FXS). In conclusion, for long tandem repeats, above ~500 bp, optical genome mapping provides an efficient method to identity repeat lengths across multiple loci simultaneously. With long intact molecules spanning repeats even kilobases in size, mosaic expansions or somatic instability can be detected with high confidence.

PrgmNr 3888 - Residual Proteome-wide Association Study Identifies Genes for Blood-Related Traits

View session detail

Author Block: **Z. Lu**[1], J. Jung[2], S. Gazal[3], N. Mancuso[1]; [1]Univ. of Southern California, Los Angeles, CA, [23]USC, Los Angeles, CA
Disclosure Block: **Z. Lu:** None.
Transcriptome-wide Association Studies (TWAS), are a recent approach to identify genes whose expression levels are associated with complex traits in large-scale GWAS data. TWAS operates by predicting mRNA levels from expression quantitative trait loci (eQTL) panels using cis-regulatory variation into large-scale GWAS cohorts. Recent studies have shown that 60% of protein QTLs (pQTL) do not colocalize with eQTLs, suggesting independent post-transcriptional/translational regulatory mechanisms which may have separate results for complex traits.
To identify genes whose total and eQTL-independent protein levels are associated with complex traits, here we describe an approach that integrates pQTL and eQTL panels with large scale GWAS summary statistics to perform a residual Proteome-wide Association Study (rPWAS), defined as a Proteome-wide Association study (PWAS) adjusted for eQTL-specific effects. We show using extensive simulations and theory that approaches to identify rPWAS genes computed directly from TWAS and PWAS summary statistics exhibit biases when eQTL partially mediates pQTL effects. To attenuate this bias, we propose an adjustment during rPWAS testing.
Having validated our approach in simulations, we integrated plasma proteomics from the INTERVAL study (N=3301), whole blood transcriptomics from GTEx (N=574), together with 27 blood-traits from the UK Biobank (mean N=350470) and performed cis-SNP heritability analyses (cis-h2g), TWAS, PWAS, and rPWAS. Of the 1139 genes matched in GTEx, we estimated an average cis-h2g of 0.042 in eQTL-residualized protein levels compared with 0.052 in total levels showing independent pQTLs explain a large portion of protein variation. Integrating total and residualized protein levels with 27 blood-traits, we identified 11122, 1448, and 472 gene-trait associations for TWAS, PWAS, and rPWAS, respectively. We found rPWAS associations were enriched for GWAS signals independent of PWAS and TWAS (P Next, we computed an enrichment score that quantifies the extent that GWAS risk is explained by predicted mRNA, protein, and residual-protein levels compared with grand association signals. We found TWAS signals to be most enriched for association signal (z = 105.47), followed by PWAS (34.54), and rPWAS (17.88), suggesting eQTL-dependent regulation is a primary mediator underlying GWAS signals.
Overall, our work sheds light onto gene regulatory mechanisms underlying blood phenotype architecture.

PrgmNr 3889 - Searching for consistent associations with a multi-environment knockoff filter

View session detail

Author Block: **M. Sesia**[1], S. Li[2], Y. Romano[3], E. CandÃ¨s[2], C. Sabatti[2]; [1]Univ. of Southern California, Los Angeles, CA, [2]Stanford Univ., Stanford, CA, [3]Technion - Israel Inst. of Technology, Haifa, Israel
Disclosure Block: **M. Sesia:** None.
This work develops a multivariate statistical method for the analysis of GWAS data that allows one to precisely discover which genetic loci are consistently associated to a certain polygenic phenotype across human populations with different ancestries. One reason why this problem is interesting is that consistency across different populations mitigates the risk of reporting spurious associations due the presence of unmeasured variants, thereby facilitating the discovery of causal associations that may be used to build more robust predictive models of genetic risk. The proposed method expands upon the existing model-X methodology of *KnockoffGWAS*, inheriting the ability of the latter to control the false discovery rate while rigorously accounting for linkage disequilibrium, population structure, and cryptic relatedness. These results are achieved without any parametric assumptions about the distribution of the phenotype, which may be either a continuous trait or a discrete disease status, for example. Numerical experiments with realistic genotype data and simulated phenotypes confirm empirically the validity of this approach. Applications to the analysis of several phenotypes in the UK Biobank data set lead to the discoveries of several consistent associations that are largely validated by other studies.

PrgmNr 3890 - Shared Genomic Segment Analysis in a Large High-Risk Chronic Lymphocytic Leukemia Pedigree Implicates *CXCR4* in Inherited Risk

View session detail

Author Block: **J. E. Feusier**, M. J. Madsen, B. J. Avery, J. A. Williams, D. M. Stephens, B. Hu, A. E. G. Osman, M. J. Glenn, N. J. Camp; Univ. of Utah, Salt Lake City, UT
Disclosure Block: **J.E. Feusier:** None.
Aim: Chronic lymphocytic leukemia (CLL) has been shown to cluster in families. First-degree relatives of individuals with CLL have an ~8 fold increased risk of developing the malignancy. Strong heritability suggests pedigree studies will have good power to localize pathogenic genes. However, CLL is relatively rare and heterogeneous, complicating ascertainment and analyses. Our goal was to identify CLL risk loci using unique resources available in Utah and methods to address intra-familial heterogeneity.**Methods**: We identified a six-generation high-risk CLL pedigree using the Utah Population Database. This pedigree contains 24 CLL cases connected by a common ancestor. We ascertained and genotyped eight CLL cases using a high-density SNP array, and then performed shared genomic segment (SGS) analysis - a method designed for extended high-risk pedigrees that accounts for heterogeneity.**Results**: We identified a genome-wide significant region (P = 1.9 Ã 10-7, LOD-equivalent 5.6) at 2q22.1. The 0.9 Mb region was inherited through 26 meioses and shared by seven of the eight genotyped cases. It sits within a ~6.25 Mb locus identified in a previous linkage study of 206 small CLL families. Our narrow region intersects two genes, including *CXCR4* which is highly expressed in CLL cells and implicated in maintenance and progression.**Conclusion**: SGS analysis of an extended high-risk CLL pedigree identified the most significant evidence to-date for a 0.9 Mb CLL disease locus at 2q22.1, harboring *CXCR4*. This discovery contributes to a growing literature implicating *CXCR4* in inherited risk to CLL. Investigation of the segregating haplotype in the pedigree will be valuable for elucidating risk variant(s).

PrgmNr 3891 - The Dynamic and Heritable Epigenetic Landscape

View session detail

Author Block: **S. Mohazzab-Hosseinian**[1], C. Breton[1], D. C. Thomas[2], D. Weisenberger[1]; [1]Univ. of Southern California, Los Angeles, CA, [2]Univ of Southern California, Los Angeles, CA
Disclosure Block: **S. Mohazzab-Hosseinian:** None.
Title: The Dynamic and Heritable Epigenetic Landscape Introduction DNA methylation is the most extensively studied epigenetic mark and is associated with many environmental exposures. A source of variability that is not well understood is multi-generational DNA methylation heritability: the stability of DNA methylation across generations that is independent of the underlying DNA sequence. Moreover, there are limited numbers of studies investigating the effect of intra-individual variation in DNA methylation. This study aims to explore heritable and dynamic epigenetic changes in a sample of 666 individuals and 732 buccal samples from 203 families in the Southern California Childrenâ□□s Health study (CHS). Among the 203 families, 66 have DNA methylation across three generations. The remaining have data on child-parent trios. Unrelated individuals with samples across two timepoints (N=66), in childhood and adulthood, were used to explore intra-individual variation and to generate methylation quantitative trait loci (mQTLs). **Methods** After quality control and normalization of the HumanMethylation EPIC array (EPIC) data, linear mixed effects models with a random effect for family or individual were used to extract intraclass correlation coefficients (ICC) for each locus. The ICC is interpreted as the variability in DNA methylation due to family or intra-individual structure. The ICC falls between zero and one, with increasing values indicating greater stability. Stability of each locus was explored across mQTLs and sequence-independent loci. Exploration of conserved versus non-conserved mQTLs across the life course was also explored. **Results** Less than 1% of features, measured by the HumanMethylation EPIC array (EPIC), were potentially heritable (ICC > 0.5). Approximately 3% of intra-individual variation was stable (ICC > 0.5). mQTL analysis indicated more significant sites in adult (N=15,709) versus child (n=8,524) samples. However, the distribution of conserved versus non-conserved mQTL loci across the life course did not vary in terms of variant consequence, genomic region, genomic feature type, epigenetic region, and epigenetic feature type. **Conclusions**Heritability from sequence independent DNA methylation loci in buccal cell samples is low. Intra-individual stability is also low. Effects of aging and the environment may contribute more variation to the epigenetic landscape than these factors.

PrgmNr 3892 - Transethnic genetic-correlation estimation and partitioning

View session detail

Author Block: **Y. Wu**, S. Sankararaman; UCLA, Los Angeles, CA
Disclosure Block: **Y. Wu:** None.
The increasing number of genetic association studies in different populations has implied the discrepancies in the genetic architecture of complex traits in different populations. Furthermore, the previous studies of non-transferability of polygenic risk scores across populations also suggest there is population-specific genetic architecture. Transethnic genetic correlation, i.e, the correlation of the effect sizes of a trait across a pair of populations, is defined in order to quantify these differences in genetic architecture. Similarly, transethnic genetic-impact correlation is another parameter defined that corrects for population-specific allele-normalization of SNP effect sizes. Additionally, it is of interest to further pin down the genetic architecture differences into smaller regions of the genome. While several methods for estimating the genome-wide transethnic genetic correlation have been proposed, there is no existing method that can estimate genome-wide transethnic genetic correlation as well as partition transethnic genetic correlation.
We propose a scalable randomized Method-of-Moments (MoM) estimator, which estimates genome-wide and partitioned transethnic genetic and genetic-impact correlation efficiently and simultaneously. This method leverages the structure of genotype data to obtain runtimes that scale sub-linearly with the number of individuals in the input dataset (assuming the number of SNPs is held constant). First, we perform extensive simulations to validate the accuracy of the estimated genome-wide transethnic genetic correlation. This method is scalable and can perform the analysis on the UK biobank dataset consisting of 430,000 individuals and 460,000 SNPs in 3 hours on a stand-alone compute machine, and partition the transethnic genetic correlation with no additional cost. We applied our method within the UK biobank on White British, Irish, and South Asian, and found novel significant transethnic genetic correlations. For example, we found the transethnic genetic correlation of HDL is 0.74(0.11) and LDL being 0.37(0.08) between White British and South Asian.

PrgmNr 3893 - Using adopted singletons to partition maternal genetic effects into pre- and post-natal effects on offspring phenotype

View session detail

Author Block: **L-D. Hwang**[1,2], G-H. Moen[1,2,3,4,5], D. M. Evans[1,2,6]; [1]Inst. for Molecular BioSci., The Univ. of Queensland, Brisbane, Australia, [2]The Univ. of Queensland Diamantina Inst., The Univ. of Queensland, Brisbane, Australia, [3]Inst. of Clinical Med., Faculty of Med., Univ. of Oslo, Oslo, Norway, [4]Population Hlth.Sci., Bristol Med. Sch., Univ. of Bristol, Bristol, United Kingdom, [5]K.G. Jebsen Ctr. for Genetic Epidemiology, Dept. of Publ. Hlth.and Nursing, NTNU, Norwegian Univ. of Sci. and Technology, Trondheim, Norway, [6]Med. Res. Council Integrative Epidemiology Unit at the Univ. of Bristol, Bristol, United Kingdom

Disclosure Block: **L. Hwang:** None.

Maternal genetic effects can be defined as the effect of a motherâ□□s genotype on the phenotype of her offspring, independent of the offspringâ□□s genotype. Maternal genetic effects can act via the intrauterine environment during pregnancy and/or via the post-natal environment. Here we develop a new model using structural equation modelling (SEM) to partition maternal genetic effects into pre- and post-natal effects. We assume that in biological families, offspring phenotypes are influenced prenatally by their motherâ□□s genotype and postnatally by their parentsâ□□ genotypes, whereas adopted individualsâ□□ phenotypes are influenced prenatally by their biological motherâ□□s genotype and postnatally by their adoptive parentsâ□□ genotypes. Critically, SEM allows us to model unobserved genotypes of the biological and adoptive parents of the adopted individuals as latent variables, permitting us to leverage the thousands of adopted singletons in the UK Biobank. We examine the utility and power of our model using simulations and asymptotic power calculations. We apply our model to educational attainment and birth weight, in up to 5178 adopted individuals, 983 trios, 3650 mother-offspring pairs, 1665 father-offspring pairs and 350330 singletons from the UK Biobank. Our results show expected patterns of maternal effects on offspring birth weight, but unexpected large prenatal effects on offspring educational attainment. Sensitivity analyses suggest this result may be due to adopted individuals in the UK Biobank being fostered by their relatives. We conclude our model can be used to estimate pre-natal and post-natal maternal genetic effects, which has exciting implications for developing pleiotropy-robust Mendelian randomization approaches for investigating the effects of maternal exposures on offspring outcomes.

PrgmNr 3894 - A benchmarking study of SARS-CoV-2 whole-genome sequencing protocols using COVID-19 patient samples

View session detail

Author Block: **Z. Chen**[1], T. Liu[1], W. Chen[1], X. Chen[1], M. Hosseini[1], Z. Yang[2], J. Li[3], D. Ho[1], D. Turay[4], C. P. Gheorghe[5], W. Jones[6], C. Wang[1]; [1]Ctr. for Genomics, Sch. of Med., Loma Linda Univ., Loma Linda, CA, [2]State Key Lab. of Respiratory Disease, Guangzhou, China, [3]State Key Lab. of Respiratory Disease, Guanzhou, China, [4]Dept. of Surgery, Sch. of Med., Loma Linda Univ., Loma Linda, CA, [5]Dept. of Gynecology & Obstetrics, Sch. of Med., Loma Linda Univ., Loma Linda, CA, [6]EA Genomics, Div. of Q2 Solutions, Morrisville, NC

Disclosure Block: **Z. Chen:** None.

The COVID-19 pandemic is a once-in-a-lifetime event, exceeding mortality rates of the flu pandemics from the 1950â□□s and 1960â□□s. Whole-genome sequencing (WGS) of SARS-CoV-2 plays a critical role in understanding the disease, particularly on the identification of newly emerging variants. Performance variation exists across SARS-CoV-2 viral WGS technologies, but there is currently no benchmarking study comparing different WGS sequencing protocols. We compared seven different SARS-CoV-2 WGS library protocols using RNA from patient nasopharyngeal swab samples under two storage conditions. We constructed multiple WGS libraries encompassing three different viral inputs: 1,000,000, 250,000 and 1,000 copies. Libraries were sequenced using two distinct platforms with varying sequencing depths and read lengths. We found large differences in mappability and genome coverage, and variations in sensitivity, reproducibility and precision of single-nucleotide variant calling across different protocols. We ranked the performance of protocols based on six different metrics. Our results indicated that the most appropriate protocol depended on viral input amount and sequencing depth. Our findings offer guidance in choosing appropriate WGS protocols to characterize SARS-CoV-2 and its evolution.

PrgmNr 3895 - An Epigenome-wide association study of body mass index (BMI) in the Multiethnic Cohort Study

View session detail

Author Block: **J. Zhu**[1], Y. M. Patel[2], S. Murphy[3], M. Tiirikainen[4], D. O. Stram[5], L. Le Marchand[6], S. L. Park[1]; [1]Univ. of Hawaii, Cancer Ctr., Honolulu, HI, [2]Univ. of Southern California, Los Angeles, CA, [3]Univ. of Minnesota, Minneapolis, MN, [4]UH Cancer Ctr., Honolulu, HI, [5]Univ of Southern California, Los Angeles, CA, [6]Univ Hawaii, Honolulu, HI
Disclosure Block: **J. Zhu:** None.
Background Epigenome-wide association studies (EWAS) of blood leukocytes have identified numerous CpG sites associated with body mass index (BMI). However, many of these studies were conducted in populations of European ancestry. The observation that BMI varies across populations and differentially influences obesity-related disease risk warrants the study of DNA methylation and BMI across racially diverse populations. We conducted an EWAS to investigate the association of DNA methylation of blood leukocytes with BMI in five racial/ethnic groups involving a total of 1,996 smokers from the Multiethnic Cohort Study (MEC). **Methods** DNA methylation of blood leukocytes were quantified from 364 African Americans, 398 Whites, 311 Native Hawaiians, 523 Japanese Americans and 400 Latinos. BMI (kg/m^2) was self-reported at time of blood draw and treated as a continuous variable. Methylation levels were assessed as a beta-value, the ratio of methylated to combined intensity at that CpG site. Linear multivariable regression analyses was conducted for all 1,996 participants for trans-ethnic estimates and for each racial/ethnic separately, with adjustment for age, sex, genetic ancestry, and cell-type distribution. Likelihood ratio test was used to test the heterogeneity of effects by race/ethnicity. To account for potential confounding by smoking dose, given that all participants were current smokers at time of blood draw, models were additionally adjusted for the urinary biomarker total nicotine equivalents (TNE). **Results** Of the 794,783 CpG sites evaluated, 670 trans-ethnic signals were found associated with BMI after Bonferroni-correction. The strongest signal was found in *ABCG1* (cg06500161, P=7.01Ã□10^{-27}), followed by *RARA* (cg17739917, P=2.28Ã□10^{-19}) and *SRRBF1* (cg11024682, P=6.21Ã□10^{-16}). These sites were all found to have an increase in methylation for increasing BMI. Our findings remained even after further adjustment for TNE, suggesting that they are independent of smoking dose. Among the 670 significantly associated CpG sites, 90 sites were found globally heterogeneous (P-interaction Conclusion This EWAS in a multiethnic population of smokers identified a great number of BMI-associated loci, many of which were also reported in previously published EWAS. Our findings suggest that while many BMI-related differentially methylated CpG sites have trans-ethnic effects, heterogeneity of the associations between DNA methylation and BMI across race/ethnicity are present and may provide mechanistic insight into observed differences in obesity-related disease risk. Replication of study findings are in progress.

PrgmNr 3896 - Evidence of causality between type 2 diabetes and dementia in the Million Veteran Program

View session detail

Author Block: **E. Litkowski**[1], M. W. Logue[2], R. Zhang[2], B. R. Charest[3], E. M. Lange[4], J. Hokanson[5], J. A. Lynch[6,7], M. Vujkovic[8], L. S. Phillips[3], R. L. Hauger[9,10], L. A. Lange[4], S. Raghavan[1], on behalf of the VA Million Veteran Program; [1]VA Eastern Colorado Hlth.care System, Aurora, CO, [2]VA Boston Hlth.care System, Boston, MA, [3]Atlanta VA Med. Ctr., Decatur, GA, [4]Dept. of Med., Univ Colorado Anschutz Med. Campus, Aurora, CO, [5]Dept. of Epidemiology, Colorado Sch. of Publ. Hlth., Univ. of Colorado, Aurora, CO, [6]Salt Lake City VA, VA Informatics & Computing Infrastructure, Salt Lake City, UT, [7]Univ. of Utah, Sch. of Med., Salt Lake City, UT, [8]Corporal Michael J. Crescenz VA Med. Ctr., Philadelphia, PA, [9]VA San Diego Hlth.care System, San Diego, CA, [10]Ctr. for Behavioral Genetics of Aging, Sch. of Med., Univ. of California San Diego, La Jolla, CA

Disclosure Block: **E. Litkowski:** None.

Background: Diabetes and dementia are diseases of high healthcare burden worldwide, and individuals with diabetes have 1.4 to 2.2 higher risk of dementia. Given previous work in the Million Veteran Program (MVP) demonstrating that a genetic risk score (GRS) for type 2 diabetes (T2D) is associated with dementia (an association that disappears when adjusting for T2D case status), our goal was to evaluate causality in the diabetes-dementia association. **Methods:** We used a two stage least squares Mendelian Randomization (MR) approach in MVP to estimate the causal associations between T2D and four types of dementia. In stage 1, we regressed a standardized, weighted genetic risk score of 331 published variants and effect sizes (GRS331) from Genome-Wide Association Studies (GWAS) against the log odds of T2D in the MVP biobank. In stage 2, we evaluated the association of the standardized genetically predicted probability of T2D from stage 1 with four nested definitions of dementia: strict Alzheimerâ□□s Disease (AD), AD plus non-specific dementias (AD-Plus), AD-related dementia (ADRD), and all-cause dementia. We used logistic regression to assess associations stratified by European (EUR, 61,986/210,469 cases/controls), African (AFR, 14,354/28440 cases/controls), and Hispanic (HIS, 6,200/11,850 cases/controls) ancestries, using a significance threshold of 0.0033 to account for multiple testing. **Result**: GRS331 predicted T2D in all three ancestries (p= 5.73e-300 to 1.54e-117). In EUR, the genetically predicted probability of T2D was associated with all dementia subtypes except strict AD (p = 4.60e-10 to 4.50e-12, OR = 1.05 to 1.06). In EUR, compared to the lowest decile of the genetically predicted probability of T2D, those in the top decile had higher odds for all dementia subtypes except strict AD (p = 0.001 to p = 5.02e-12, OR = 1.18 to 1.31). There were no associations between the genetically predicted probability of T2D and dementia in AFR and HIS. **Conclusion:** We found evidence of causality between diabetes and dementia in EUR. A limitation of this study is the lack of statistical power in AFR and HIS. Further analysis is required to systematically evaluate pleiotropy in all ancestries.

PrgmNr 3897 - Examining genetic associations with liver steatosis in Mexican-origin adults

View session detail

Author Block: **M. Trejo**, K. E. Morrill, Y. C. Klimentidis, D. O. Garcia; Univ. of Arizona, Tucson, AZ
Disclosure Block: **M. Trejo:** None.
Background: Various studies have identified single-nucleotide polymorphisms (SNPs) related to non-alcoholic fatty liver disease (NAFLD), specifically ones located in or near the *LYPLAL1*, *GCKR*, *PPP1R3B*, *TM6SF2*, *MBOAT7*, and *PNPLA3* genes. However, these SNPs were identified primarily in populations of European ancestry. This study examined the associations of these previously identified SNPs with liver steatosis in a sample of Mexican-origin adults living in Southern Arizona. **Methods:** A total of 307 Mexican-origin adults between the ages of 18 and 64 with a body mass index (BMI) of 25 kg/m^2 or higher were genotyped at the following SNPs: rs12137855 (*LYPLAL1*), rs1260326 (*GCKR*), rs4240624 (*PPP1R3B*), rs58542926 (*TM6SF2*) rs641738 (*MBOAT7)* and rs738409 (*PNPLA3*). All had liver steatosis assessed through transient elastography (FibroScanÂ®). Additive, dominant, and recessive regression models examined the association between the six SNPs and liver steatosis. Age and BMI were examined as potential modifiers of genetic associations. **Results:** Participants were, on average, 45 years old and mostly female (63%) with an overall mean liver steatosis of 288.1 dB/m, indicative of steatosis >5%. Models showed no association between *LYPLAL1*, *GCKR*, *PPP1R3B*, *TM6SF2*, or *MBOAT7* and liver steatosis. Only *PNPLA3* was statistically significantly associated with liver steatosis in both additive and recessive models (pConclusion: SNPs associated with NAFLD in populations of European descent did not strongly contribute to liver steatosis in individuals of Mexican-origin, except for rs738409 (*PNPLA3*). Further efforts are necessary to explore additional SNPs that may be associated with NAFLD in this high-risk population.

PrgmNr 3898 - Genetic determinants of prostate-specific antigen levels improve cancer screening utility

View session detail

Author Block: **L. Kachuri**[1], R. E. Graff[1], J. P. Shelley[2], K. Schaffer[2], J. D. Mosley[2], P. J. Goodman[3], C. Till[3], I. Thompson[4], R. Klein[5], S. K. Van Den Eeden[6], T. J. Hoffmann[1], J. S. Witte[7]; [1]Univ. of California San Francisco, San Francisco, CA, [2]Vanderbilt Univ. Med. Ctr., Nashville, TN, [3]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [4]Christus Hlth., San Antonio, TX, [5]Icahn Sch. of Med. at Mt. Sinai, New York, NY, [6]Kaiser Permanente Northern California, Oakland, CA, [7]Univ California San Francisco, San Francisco, CA

Disclosure Block: **L. Kachuri:** None.

Prostate-specific antigen (PSA) testing is controversial due to issues related to sensitivity and specificity, resulting in overdiagnosis and overtreatment of prostate cancer (PCa). Genetic determinants of PSA in cancer-free men could be used to correct observed PSA values by accounting for PSA variation that does not reflect PCa.

We conducted the largest ever genome-wide association study (GWAS) of PSA in men without PCa (N=65,962; 63,338 European ancestry) using longitudinal measures from the UK Biobank (UKB; n=26,491), BioVU (n=8078), and Genetic Epidemiology Research on Adult Health and Aging (GERA; N=30,088) cohorts. Our GWAS discovered 87 variants associated with PSA levels (P-8), including 37 loci independent of previously reported associations (r^2PSA) constructed from these variants was validated in the Prostate Cancer Prevention Trial (PCPT; P=8.7Ã□10^{-54}). In the PCPT, which enrolled PCa-free men with PSAâ□¤3 ng/mL, PRS explained a larger proportion of PSA variation than age (4.0% vs. 1.3%).

Consistent with the hypothesis that genetic predisposition to elevated PSA increases the detection of low-grade PCa, PRS_{PSA} was inversely associated with Gleason score (â□¤6 vs. â□¥8: OR=0.83, P=8.4Ã□10^{-5}; 6485 GERA cases) and PCa mortality (HR=0.82, P=7.4Ã□10^{-9}; 8834 UKB cases). We then evaluated how genetically corrected PSA values affect reclassification at cut-offs used for biopsy recommendations in GERA. Among non-cases with a negative biopsy, 18.4% were reclassified below the referral threshold, while 2.7% moved upward. In cases, downward reclassification (3.9%) was higher than upward (1.9%) when considering PSA values â□¥2 years before diagnosis. This trend was more pronounced in cases with low-risk disease (Gleason Lastly, we explored the role of PSA-related selection bias on associations with PCa risk and mortality. There is substantial sharing of genetic loci between PSA and PCa, illustrated by the high correlation between their genetic scores (r=0.287, P-500) in UKB (n=164,669 not included in the PSA GWAS). Although this may partly reflect pleiotropy, there was evidence of bias due to screening (P=2.1Ã□10^{-130}). Re-fitting PRS_{PCa} using bias-corrected risk allele weights attenuated its correlation with PRS_{PSA} (r=0.049, P=5.5Ã□10^{-93}) and revealed a previously absent association with PCa mortality (HR=1.18, P=0.035).

Our work provides preliminary evidence that genetic correction of PSA levels may improve PCa screening. Larger and more diverse study populations are required to fully characterize the genetic basis of PSA variation and optimize its clinical utility.

PrgmNr 3899 - Genetic factors explain varying glaucoma prevalence and pressure subtype across ancestries

View session detail

Author Block: **P. Gharahkhani**[1], X. Han[2], J. ONG[3], S. Macgregor[1]; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]300 Herston Road, BRISBANE, Australia, [3]QIMR Berghofer Med. Res. Inst., HERSTON, Australia
Disclosure Block: **P. Gharahkhani:** None.
Glaucoma is the leading cause of irreversible blindness. Increases in intraocular pressure (IOP) and in optic nerve vertical cup-to-disc ratio (VCDR) are associated with greater glaucoma risk. Glaucoma has the highest prevalence in individuals of African ancestry, followed by Latinos, then Europeans and East Asians. East Asians are much more frequently diagnosed with a low pressure glaucoma subtype, but the reasons for this are hotly debated. We aimed to investigate whether variation in glaucoma presentation can be explained by differences in the distributions of polygenic risk across ancestries.We applied an artificial intelligence approach to allow accurate and uniform optic nerve phenotyping in two large independent biobanks, UK Biobank and Canadian Longitudinal Study on Aging. We then derived polygenic risk scores (PRS) for glaucoma, IOP and VCDR and applied them to a large independent multi-ethnic cohort (Kaiser Permanente). Within Kaiser, Africans had greatly elevated (~1 standard deviation higher) glaucoma PRSs, followed by Latinos, Europeans and East Asians, mirroring reported prevalences. The ordering of the ancestries was similar for the IOP PRS, mirroring observations that Africans have high glaucoma rates (particularly high pressure glaucoma), whilst Asians are affected at lower rates (with the low pressure subtype common). The VCDR PRS was higher in East Asians/Africans than in other ancestries. After adjusting for the IOP and VCDR PRSs, glaucoma PRS means were indistinguishable across ancestries, suggesting the observed ancestry differences are primarily due to genetic effects on IOP and VCDR. The results were essentially unchanged in our sensitivity analyses investigating population structure biases.Together, our findings indicate that, relative to Europeans, African ancestry individuals have both genetically elevated VCDR and IOP, explaining the elevated prevalence of glaucoma. By contrast, a combination of genetically lower IOP but elevated VCDR may explain the similar glaucoma prevalence in Europeans and East Asians, and suggesting a genetic explanation for the â□□low tensionâ□□ form of glaucoma common in East Asia. These findings are consistent with our findings at the phenotypic level (Han et al., AJHG 2021), suggesting there are genetic drivers behind the different prevalence and types of glaucoma which manifest in different ancestries.

PrgmNr 3900 - Heritability of audiometric phenotypes using multigenerational pedigrees

View session detail

Author Block: **J. Duran**[1], S. Mansour[1], H. Coon[2], L. B. Jorde[3]; [1]Dept Human Gen, Univ of Utah, Salt Lake City, UT, [2]Univ. of Utah Sch. of Med., Salt Lake City, UT, [3]Univ of Utah, Salt Lake City, UT
Disclosure Block: **J. Duran:** None.
Little is known about the heritability of audiometric phenotypes. Current heritability estimates are based on twin studies, which often yield upwardly biased estimates due to violations of assumptions about shared environment and gene-environment interactions. We have employed the Sequential Oligogenic Linkage Analysis Routines (SOLAR) algorithm to estimate the heritability of hearing phenotypes in the Utah CEPH pedigrees. SOLAR uses pedigree data intersected with phenotype and genotype data to build identity by decent matrices to estimate heritability. Additionally, SOLAR uses maximum likelihood methods to decide whether a trait has a significant genetic component and whether the mode of inheritance is Mendelian or polygenic. Previous twin studies of hearing acuity provide a wide range of heritability estimates: 45 to 75 percent for standard audiometric measures and an even larger range of variation for hearing loss. We hypothesize that twin studies overestimate the heritability of hearing acuity and that using three-generational pedigree data will produce a more accurate estimate of the heritability of hearing acuity. We quantified hearing acuity as the two-ear average of hearing thresholds at high frequency and pure-tone frequency (frequencies at which human voices are heard). Our analysis controlled for age and sex as covariates, which were not previously accounted for in twin studies. Our findings indicate that high-frequency hearing acuity has a polygenic mode of inheritance with an estimated heritability of 23 percent. Similarly, pure-tone hearing acuity has a polygenic mode of inheritance and a heritability is 27 percent. Thus, for both phenotypes, heritability estimates are decreased compared to previously reported twin studies.

PrgmNr 3901 - Methylation patterns associated with inflammation traits in racially and ethnically diverse populations

View session detail

Author Block: **J. Lundin**[1], U. Peters[2], Y. Hu[3], C. Carlson[1], C. A. Haiman[4], K. E. North[5], L. M. Raffield[6], C. Avery[7], T. C. Matise[8], S. G. Buyske[8], C. R. Gignoux[9], Z. Du[1], M. Kim[10], E. Kenny[11], R. J. F. Loos[12], A. Reiner[13], C. Kooperberg[1]; [1]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [2]Fred Hutchinson CA Res Ctr, Seattle, WA, [3]Fred Hutch, Seattle, WA, [4]Los Angeles, CA, [5]Univ North Carolina, Chapel Hill, NC, [6]UNC - Chapel Hill, Chapel Hill, NC, [7]Univ. of North Carolina, Chapel Hill, NC, [8]Rutgers Univ, Piscataway, NJ, [9]Univ of Colorado Denver, Anschutz Med. Campus, Aurora, CO, [10]Fred Hutchinson Cancer Res. Ctr., Seattle, [11]Icahn Sch. of Med. at Mt Sinai, New York, NY, [12]The Icahn Sch. of Med. at Mount Sinai, New York, NY, [13]Univ of Washington, Seattle, WA

Disclosure Block: **J. Lundin:** None.

Background. Inflammation is a complex immune response including chronic low-grade inflammation that contributes to the pathophysiology of chronic diseases such as cardiovascular disease, cancer and diabetes. Genome-wide association studies have identified candidate loci of genetic variants associated with biomarkers of inflammation clustered by immune pathways and liver metabolic pathways. However, the proportion of variance explained by genetic loci remains small and has largely been evaluated in only European Ancestry populations. In this study we extend this evaluation to the epigenome to evaluate DNA methylation associated with biomarkers of inflammation across multiple races and ethnicities. **Methods.** We meta-analyzed epigenome-wide association studies (EWAS) of inflammation biomarkers (C-reactive protein, CRP; IL-6; fibrinogen) measured in peripheral blood samples with DNA methylation levels (from whole blood assays) in the Womenâ□□s Health Initiative (WHI) cohort (discovery) along with diverse replication cohorts. We analyzed CpG sites common to the Infinium HM450K and EPIC arrays using mixed effects linear regression with the inflammation marker level as the outcome variable, and DNA methylation as the predictor, adjusted for age, sex, body-mass index (BMI), smoking (ever, former, current), population structure (via PC1-10), cell type composition (proportion of WBC species), and batch effects (as random effect). Significance was defined as pResults. The discovery stage (n=4,378 from the WHI cohort) identified 100 significant CpG sites associated with CRP values, 24 associated with fibrinogen, and 5 with IL-6 in the multi-racial/ethnic analyses. The nearby genes for 14 of the top 20 most significant DNA methylation sites associated with CRP level (pConclusion. Large-scale EWAS of inflammation biomarkers identified significant CpGs associated with CRP, fibrinogen, and IL-6. Next steps include exploring biological pathways, inflammation related disease (e.g., cardiovascular disease, diabetes), and associated medication use as potential confounders and effect modifiers, and extending these analyses to replication cohorts including ancestrally diverse populations. It is currently unknown whether the identified CpG-lipid associations could be generalized to other racial/ethnic groups or replicated for African Ancestry. Further analyses include performing race/ethnic-specific EWAS models to explore the generalizability and heterogeneity of inflammation marker-associated CpG sites.

PrgmNr 3902 - ML-based COPD phenotype denoising using volumetric flow data improves genomic discovery and disease risk prediction

View session detail

Author Block: **J. Cosentino**[1], Z. McCaw[1], C. McLean[2], F. Hormozdiari[2], B. Behsaz[2], B. Alipanahi[1]; [1]Google Hlth., Palo Alto, CA, [2]Google Hlth., Cambridge, MA
Disclosure Block: **J. Cosentino:** Major Stockholder/Ownership Interest; Google. Salary/Employment; Google. Receipt of Intellectual Property Rights/Patent Holder; Google.
Complex disorders such as chronic obstructive pulmonary disease (COPD) have multiple etiologies with similar manifestations, making the process of phenotyping a large cohort arduous. Even after curating a COPD phenotype using over 600 self-reported, general practitioner, and hospital records from the UK Biobank (UKB), a GWAS for 364,256 individuals identified only 5 genome-wide significant (GWS) loci, failing to replicate most known COPD-associated loci. COPDâ□□s low prevalence (~4%) and the large fraction of undiagnosed (81%) and false positive (62%) cases imply that, although sourced from medical records, these manually-defined labels are inherently noisy. Here we show that deep neural networks (DNNs) can denoise such labels, allowing us to improve genomic discovery by amplifying sparse signals found in manually-generated phenotypes.
Using our noisy labels, we trained an ensemble of 10 DNNs to predict COPD status from raw volumetric flow data, consisting of 1,000 flow-volume measurements collected at 10ms intervals. Our model achieved an AUPRC=0.27 (95% CI: 0.25-0.30; n=31,965; prevalence=3.95%), outperforming spirometry-based predictions defined using the global initiative for chronic obstructive lung disease (GOLD) criteria (AUPRC=0.19 [0.17-0.21]). We split unrelated UKB individuals of European ancestry into two folds, trained a model on each fold, and applied each model to the other fold.We then ran a GWAS using COPD risks predicted by our model on 269,405 individuals (ML-based GWAS), controlling for age, sex, height, BMI, genetic PCs, smoking status, and the data fold. The ML-based GWAS identified 215 GWS loci, replicating 39 out of 41 previously reported COPD loci (Sakornsakolpat et al., Nat. Genet. 2019). The top locus is *4q24*, encompassing *NPNT* (rs34712979, P=9.9e-106), a known COPD susceptibility locus. Functional enrichments of the GWS loci highlight development and morphogenesis of respiratory (3) and cardiac (11) terms, consistent with known cardiac comorbidities of COPD. A GWAS using the GOLD criteria phenotype identified only 17 GWS loci, 16 of which were also identified by the ML-based GWAS, and no functional enrichments.
To validate these GWASs on an independent cohort, we computed polygenic risk scores (PRSs) for the COPDGene dataset, a clinically curated cohort of ~10k individuals. The ML-based and GOLD PRSs resulted in AUPRCs of 0.61 (0.60-0.63; prevalence=52.7%) and 0.57 (0.55-0.59) on 5,269 individuals of European ancestry, respectively. This work illustrates the potential for using ML to improve genomic discovery by amplifying signals from otherwise noisy phenotypes into quantitative risk scores.

PrgmNr 3903 - Multitrait and AI-enhanced GWAS of glaucoma and its risk factors enables PRS based prediction of risk which is portable across ancestries

View session detail

Author Block: **S. Macgregor**[1], Y. Wu[1], X. Han[1], P. Gharahkhani[1], A. W. Hewitt[2], J. E. Craig[3], D. A. Mackey[4], J. L. Wiggs[5], A. P. Khawaja[6], M. Trzaskowski[7]; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]Ctr Eye Res Aust, Brunswick, Australia, [3]Flinders Univ., Adelaide, Australia, [4]Lions Eye Inst., Perth, Australia, [5]Harvard Med Sch, MEEI, Boston, MA, [6]Moorfields Eye Hosp, London, United Kingdom, [7]Max Kelsen, Brisbane, Australia
Disclosure Block: **S. Macgregor:** Major Stockholder/Ownership Interest; I hold stock in StratifEYE Pty Ltd. Receipt of Intellectual Property Rights/Patent Holder; I am an author of a filed glaucoma genetic risk score patent.
Glaucoma, one of the most heritable human diseases, is the leading cause of irreversible blindness. Glaucoma is usually treatable but only if diagnosed in time. Previous studies in European ancestry populations have shown that polygenic risk scores (PRS) offer excellent potential for stratifying people into glaucoma risk categories. Glaucoma incidence is highest in those of African ancestry, followed by Asians, then Europeans. Alas, genetic studies to date have focused on the latter two, particularly Europeans.
Some early studies suggested that the genes underlying glaucoma risk differ across ancestries. However, the latest cross-ancestry GWAS showed there is substantial concordance in the effect sizes at implicated risk loci, suggesting that PRS from one population will have utility in the others.We performed a cross ancestry multitrait GWAS harnessing data from both glaucoma cohorts (International Glaucoma Genetics Consortium) and glaucoma endophenotypes, with AI based phenotyping used to maximize precision. We built a PRS and tested its predictive performance in independent cohorts of European, South Asian, East Asian and African ancestry.
The multitrait GWAS greatly increased the effective sample size, resulting in increased prediction accuracy relative to our 2020 PRS in all ancestries (e.g. AUC increases over non-genetic baseline 0.05 and 0.07 for 2020 and 2021 PRS, respectively in Europeans; analogous figures in Africans 0.03 and 0.06). As expected given the Euro-centric training set, PRS performance in non-Europeans was typically reduced relative to Europeans but AUC increases were significant relative to baseline. In Africans those in the top PRS decile had >7 fold increased risk relative to the bottom decile. In South Asians, those in the top PRS decile reached 3% glaucoma prevalence 18 years earlier than those in the bottom decile.
Future GWAS should better represent all ancestry groups to ensure good PRS performance worldwide. In the meantime, our current cross-ancestry GWAS enables construction of a PRS which performs acceptably in all major ancestries, including African ancestry populations where glaucoma rates are very high. Our new glaucoma PRS has excellent potential to help prevent glaucoma blindness worldwide.

PrgmNr 3904 - Multi-trait Genome-wide association study identifies variants associated with corneal parameters that contributing to keratoconus risk

View session detail

Author Block: **W. HE**[1], X. Han[2], J. ONG[3], P. Gharahkhani[1], S. Macgregor[1]; [1]QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]300 Herston Road, BRISBANE, Australia, [3]QIMR Berghofer Med. Res. Inst., HERSTON, Australia

Disclosure Block: **W. He:** None.

Keratoconus (KC) is a condition in which the cornea thins, bulges, and flattens into a conical shape. Sequelae of KC include myopia and astigmatism. KC is the main cause of corneal transplantation, with a prevalence of 1.38 per 1000 worldwide. Obtaining large KC datasets for GWAS is problematic because of the relative rarity of KC. However, recent large studies have measured quantitative corneal parameters and investigating them may provide insight into the pathogenesis of KC. Central corneal thickness (CCT) is a corneal parameter directly related to the corneal thinning character of KC. Another parameter that has been found to be significantly lower in keratoconus patients is the corneal resistance factor (CRF). Many large-scale cohorts with genetic data have collected CCT or CRF data, but not both. Here, we performed a multi-trait GWAS with European ancestry CRF data from the UK Biobank and the Canadian Longitudinal Study on Aging (CLSA) (N_{UKB}=105427, N_{CLSA}=18190), and European ancestry CCT data from International Glaucoma Genetics Consortium (IGGC) (N=17803). We identified 383 CRF and 238 CCT loci, including 171 CRF and 176 CCT loci that had not previously been reported. We replicated our result in UK Biobank and IGGC Asian ancestry groups. The power to replicate individual loci was low but overall they displayed a strong concordance in effect size across ancestries, indicating the possible cross-ancestry effects within CRF and CCT loci. We looked up the novel CRF/CCT-associated loci (160 CRF and 161 CCT) in a recent KC GWAS (4669 cases, 116547 controls). As expected many of the CCT/CRF loci were strongly associated with KC (P CCDC80, FBXO31, HERC2, CD34 and MAMDC2. We found negative correlations in effect size between two corneal parameters and KC. A two-sample Mendelian randomisation analysis revealed that CRF and CCT are likely to increase the risk of KC causally. These findings demonstrate the power of multi-trait GWAS using corneal parameters to identify new KC risk loci.

PrgmNr 3905 - Perceived aging in the UK Biobank - A proxy measure for skin aging?

View session detail

Author Block: **N. Ingold**[1,2], X. Han[1], C. M. Olsen[3,4], D. C. Whiteman[3,4], J. Ong[1], S. Macgregor[5], M. H. Law[1,2]; [1]Statistical Genetics, QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [2]Sch. of BioMed. Sci., Faculty of Hlth., Queensland Univ. of Technology, Brisbane, Australia, [3]Dept. of Population Hlth., QIMR Berghofer Med. Res. Inst., Brisbane, Australia, [4]Faculty of Med., The Univ. of Queensland, Herston, Queensland, Australia, [5]QIMR Berghofer Med. Res. Inst., Brisbane, Australia
Disclosure Block: **N. Ingold:** None.
Age is perceived through the presentation of a range of features on the skin such as wrinkles. These features can arise from sources both intrinsic (e.g. physiological changes over time) or extrinsic (e.g. sun exposure). Some components of skin aging are also associated with risk of skin cancer. To date, the largest genome-wide-association study (GWAS) of perceived (relative) age using the UK Biobank (N=~400,000) identified 74 genetic loci. We proposed two aims; 1) to leverage power from genetically correlated phenotypes to identify additional loci associated with perceived aging, and 2) to evaluate the overlap between the genetics of perceived age and facets of skin aging relevant to skin cancer. Genetic correlations (rg) between perceived younger age and multiple traits were assessed using LD score regression (LDSC), both generally and by sex. LDSC indicated that perceived age in men and women may have differing genetic architecture (female vs. male rg = 0.75, 95% CI = 0.69 to 0.80). Higher BMI (rg = -0.2, 95% CI = -0.24 to -0.17) and greater cigarettes per day (rg=-0.15, 95% CI=-0.21 to -0.10) were genetically associated with older perceived age in both sexes, while baldness was correlated with greater perceived age in males(rg = -0.36, 95% CI = -0.41 to -0.30) and pigmentation traits were correlated with looking younger in females (rg = 0.21, 95% CI = 0.10 to 0.33). Phenotypes that directly measure components of skin aging were not correlated with perceived age (rg = 0.02, 95% CI = -0.19 to 0.22), nor was melanoma (rg = -0.01, 95% CI = -0.12 to 0.08). We will report on using multi-trait analysis of GWAS summary statistics (MTAG) to identify additional genes associated with perceived age, and to further explore aim 2 we will apply GWAS subtraction to attempt to identify components of perceived age (e.g. extrinsic skin aging). We hypothesised that additional loci can be identified for perceived age in the UK-Biobank using multi-trait analysis, and that perceived age could also provide a measure of components of skin aging relevant to skin cancer risk. While the results do not reflect perceived age as being strongly driven by skin aging, our analysis is a rare example of a complex trait with differences in genetic architecture by sex.

PrgmNr 3906 - Quantifying uncertainty and variability of recessive disease prevalence using Monte Carlo estimates

View session detail

Author Block: **S-G. Ji**, P. Komar, K. Shah, M. Soliai, C. Barnhill; BridgeBio Pharma, Palo Alto, CA
Disclosure Block: **S. Ji:** Salary/Employment; BridgeBio Pharma.

Disease prevalence is a key metric of biomedical research and therapeutic development prioritization. The more accurately the number of individuals who may benefit from a potential therapy can be predicted, the better informed decisions can be made. Unknowns such as rate of underdiagnosis, and regional or ethnic variability induce significant uncertainty, sometimes on the order of 100x. Quantifying this uncertainty is necessary for informed public health decision making regarding the disease burden.

We present a statistical approach to robustly estimate the ethnicity- and region-specific incidence rate of recessive diseases based on allele count data from large population cohorts. In turn, prevalence is computed from the estimated incidence and life-expectancy. Multiple population-level data (gnomAD, UK Biobank, TOPMed), with different inclusion criteria, are used to increase the precision of the estimates. First, we limit the variants to those with pathogenic and likely pathogenic mutations based on ACMG guidelines for lower bound estimates. Second, we draw Monte Carlo Markov chain (MCMC) samples of ethnicity-specific allele frequencies from a non-linear random-effect model with binomial outcome and logit-normal prior shared between ethnic groups. Finally, allele frequency samples are combined across variants to produce ethnicity-specific frequencies of carriers, assumed to be unaffected, and aggregated to yield region-specific incidences assuming Hardy-Weinberg. For an upper limit, we repeat the approach with a relaxed variant filtering. We filter based on functional annotation from sources ranging from pLoF to multiple computational predictions. Computing with MCMC samples allows one to characterize the uncertainty and enables propagating it to downstream analyses accurately. A Python implementation completes one such analysis in a few minutes.

We validated our method by comparing its results on 24 autosomal recessive diseases for which genetic newborn screen data is available. Our method was capable of accurately predicting the incidence of both ultra-rare (1 in a million births) and rare (1 in 10,000 births) diseases. The results were concordant within a margin of factor of 3 for 21 of the 24 diseases. After validation, we successfully employed the method on Recessive Dystrophic Epidermolysis Bullosa to predict 75% higher incidence (5.4, 95%CI:[2.8, 9.6] in 1 million births) than epidemiology literature (3 in 1 million births) that was corroborated by expert review and extrapolation from claims data. This method is reliable for inherited, recessive diseases with clear genetic causes.

PrgmNr 3907 - Tissue-specific functional annotations highlight association of liver polygenic risk score with Alzheimerâ□□s disease and related biomarkers

View session detail

Author Block: **D. Panyard**[1,2], Y. K. Deming[1,3,4], B. F. Darst[5], C. A. Van Hulle[3,4], H. Zetterberg[6,7,8,9], K. Blennow[6,7], G. Kollmorgen[10], I. Suridjan[11], C. M. Carlsson[3,4,12], S. C. Johnson[3,4,12], S. Asthana[3,4,12], C. D. Engelman[13], Q. Lu[13,14]; [1]Dept. of Population Hlth.Sci., Univ. of Wisconsin-Madison, Madison, WI, [2]Dept. of Genetics, Stanford Univ., Stanford, CA, [3]Wisconsin Alzheimerâ□□s Disease Res. Ctr., Univ. of Wisconsin-Madison, Madison, WI, [4]Dept. of Med., Univ. of Wisconsin-Madison, Madison, WI, [5]Keck Sch. of Med., Univ. of Southern California, Los Angeles, CA, [6]Dept. of Psychiatry and Neurochemistry, Inst. of NeuroSci. and Physiology, the Sahlgrenska Academy at the Univ. of Gothenburg, MÃ¶lndal, Sweden, [7]Clinical Neurochemistry Lab., Sahlgrenska Univ. Hosp., MÃ¶lndal, Sweden, [8]Dept. of Neurodegenerative Disease, UCL Inst. of Neurology, London, United Kingdom, [9]UK Dementia Res. Inst. at UCL, London, United Kingdom, [10]Roche Diagnostics GmbH, Penzberg, Germany, [11]Roche Diagnostics Intl. Ltd., Rotkreuz, Switzerland, [12]William S. Middleton Mem. Veterans Hosp., Madison, WI, [13]Univ. of Wisconsin-Madison, Madison, WI, [14]Dept. of Statistics, Univ. of Wisconsin-Madison, Madison, WI
Disclosure Block: **D. Panyard:** None.

Alzheimerâ□□s disease is a neurodegenerative disease whose causal mechanisms are not yet fully known. The use of functional annotation in genetic analyses has been shown to be able to enhance polygenic risk score (PRS) models, improving their power and identifying new insights into disease mechanisms. Here, we leveraged genomic functional annotations from GenoSkyline-PLUS that provide a tissue-specific measure of which genomic regions are expected to be functional. Using these annotations, individual-level genetic data from participants of European ancestry from the Wisconsin Registry for Alzheimerâ□□s Prevention (WRAP, n = 1,198) and Wisconsin Alzheimerâ□□s Disease Research Center (WADRC, n = 212) cohorts, and genome-wide association study summary statistics from the International Genomics of Alzheimerâ□□s Project (IGAP 2019), we built tissue-specific PRS models for 13 tissues and applied the scores to two longitudinal cohort studies of AD that include both cognitive diagnoses and a rich set of cerebrospinal fluid biomarkers for AD, neurodegeneration, and inflammation measured with the Roche NeuroToolKit immunoassays. The model most strongly associated with AD diagnosis (and the only model statistically significantly associated after the strongly associated APOE locus was removed) was the liver PRS: n = 1,116; OR = 2.19 (1.70-2.82); P = 1.46×10^{-9}. This liver PRS was also statistically significantly associated with two major AD biomarkers in the cerebrospinal fluid: amyloid (P = 3.53×10^{-6}) and phosphorylated tau (P = 1.45×10^{-5}). These findings highlight the potential of functional annotation in PRS studies and provide new evidence highlighting the role of the liver-functional genome in AD.

PrgmNr 3908 - Utility of polygenic risk scores for colorectal cancer risk assessment across diverse populations

View session detail

Author Block: **M. Thomas**[1], L. C. Sakoda[2], M. Timofeeva[3], V. DÃez-Obrero[4], Z. Chen[5], P. Law[6], E. A. Rosenthal[7], S. Schmit[8], C. Fernandez-Rozadilla[9], C. R. Gignoux[10], J. C. Figueiredo[11], J. Huyghe[12], T. A. Harrison[12], R. B. Hayes[13], I. Tomlinson[14], S. B. Gruber[11], G. Casey[15], R. S. Houlston[16], V. Moreno[17], W. Zheng[18], M. Dunlop[19], G. P. Jarvik[20], D. A. Corley[21], U. Peters[22], L. Hsu[12]; [1]Fred Hutch, Seattle, WA, [2]Kaiser Permanente, Oakland, CA, [3]Univ. of Southern Denmark, Odense, Denmark, [4]Colorectal Cancer Group, Molecular Mechanisms and Experimental Therapy in Oncology (ONCOBELL) Program, Bellvitge BioMed. Res. Inst. (IDIBELL), Barcelona, Spain, [5]Div. of Epidemiology, Vanderbilt Univ. Med. Ctr., Nashville, TN, [6]Div. of Genetics and Epidemiology, The Inst. for Cancer Res., London, United Kingdom, [7]Univ. of Washington Sch. of Med., Seattle, WA, [8]Cleveland Clinic, Cleveland, OH, [9]Inst. de Investigacion Sanitaria de Santiago - IDIS, Santiago de Compostela, Spain, [10]Univ of Colorado Denver, Anschutz Med. Campus, Aurora, CO, [11]USC, Los Angeles, CA, [12]Fred Hutchinson Cancer Res. Ctr., Seattle, WA, [13]Div. of Epidemiology, Dept. of Population Hlth., New York Univ. Sch. of Med., New York, NY, [14]Cancer Res. UK Edinburgh Ctr., Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom, [15]Ctr. for Publ. Hlth.Genomics, Univ. of Virginia, Charlottesville, Virginia, VA, [16]The Inst. of Cancer Res., Sutton, United Kingdom, [17]Oncology Data Analytics Program, Catalan Inst. of Oncology-IDIBELL, L'Hosp.et de Llobregat, Barcelona, Spain, [18]Vanderbilt Univ Med Ctr., Nashville, TN, [19]Colon Cancer Genetics Group, Med. Res. Council Human Genetics Unit, Inst. of Genetics and Cancer, Univ. of Edinburgh, Edinburgh, United Kingdom, [20]Univ Washington Med Ctr., Seattle, WA, [21]Div. of Res., Kaiser Permanente Northern California, Oakland, CA, [22]Fred Hutchinson CA Res Ctr, Seattle, WA
Disclosure Block: **M. Thomas:** None.

Colorectal cancer (CRC) is a leading cause of cancer-related mortality, yet many CRC are preventable by screening individuals at increased disease risk. Polygenic risk score (PRS) offer the prospect of precision prevention by defining individualâ□□s risk of CRC and tailoring screening. Genome-wide CRC-PRS built upon European (EUR) data (PMID: 32758450) have limited performance in non-EUR populations because of small number of non-EUR populations included in the training data set. To address this deficiency, we expanded our PRS development to include non-EUR populations.

We derived PRS by leveraging GWAS summary statistics and LD structure from EUR and East Asian (EAS) ancestry. Our training data contains 78,473 cases & 107,142 controls of EUR ancestry and 21,737 cases & 47,444 controls of EAS ancestry. We examined 3 approaches for PRS development, using 1) 180 CRC known loci, 2) genome-wide EAS and EUR summary statistics and LD matrices, 3) combined genome-wide EUR and EAS summary statistics and weighted LD matrices with weights defined as the proportion of subjects from each ancestry. For Approaches 2 and 3, we used LDpred2 to derive PRS including ~1M SNPs. For Approaches 1 and 3 we derived a single PRS across EAS and EUR populations, whereas for Approach 2 we first derived ancestry-specific PRSs (PRSEUR and PRSEAS) and then final PRS for each ancestry by leveraging PRS from other ancestry using a weighted sum of PRSEUR and PRSEAS where the weights were obtained from a multivariate regression model, using 2,627 cases & 3,797 controls for EAS and 29,864 cases & 31,629 controls for EUR ancestry. To evaluate the performance, we analyzed independent data from Genetic Epidemiology Research on Adult Health and Aging cohort of 101,987 individuals, which included 1,699 CRC cases.

The AUC of the best performing approach, which combined genome-wide EUR and EAS summary statistics and LD matrices (Approach 3), were 0.67, 0.65, 0.63 and 0.57 for European, East Asian, Latinx, and African American ancestry groups, respectively, with hazard ratios of top 30% vs. remaining for corresponding groups of 2.52, 1.61, 2.15 and 1.37. Compared to EUR-based PRS (PMID: 32758450), the AUCs for Approach 3 are improved by 3%, 6%, 5% and 2% for European, East Asian,

Latinx and African American ancestry groups, respectively.
A trans-ethnic PRS has the potential to reduce disparities in PRS performance between non-EUR and EUR which is critical if PRS is going to be implemented in clinical practice or populations screening programs. As the performance remains lower in non-EUR populations, larger and more racially and ethnically diverse study populations may further improve equitable risk prediction.

PrgmNr 3909 - mTOR inhibition improves lymphatic dysplasia caused by*PIK3CA*p.H1047L in preclinical models

View session detail

Author Block: **S. Sheppard**[1], M. Battig[1], S. Seiler[1], J. Belasco[1], A. GANGULY[2], C. Kao[3], S. E. Kim[1], D. Li[1], D. W. Low[1], M. March[1], L. Matsuoka[1], R. Niazi[4], V. Sanders[1], A. S. Srinivasan[1], L. Surrey[1], D. Adams[1], H. Hakonarson[1]; [1]Children's Hospital of Philadelphia, Philadelphia, PA, [2]University of Pennsylvania, Philadelphia, PA, [3]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, [4]Genetic Diagnostic Laboratory, Department of Genetics, University of Pennsylvania, Philadelphia, PA
Disclosure Block: **S. Sheppard:** Major Stockholder/Ownership Interest; HCA. Royalty(ies)/Honoraria; UpToDate.

Lymphatic malformations (LM) are debilitating congenital lesions that arise from abnormal development of the lymphatic system. The child was born at 26 weeks due to premature rupture of membranes and required invasive ventilation for three months. After extubation, a vascular area on the tongue was noted with concern for vascular malformation. Around 9 months of age he had CO_2 laser excision of a large laryngeal cyst obstructing the airway. Histology demonstrated irregular D-240 and CD31 positive lymphatic channels, consistent with the diagnosis of lymphatic malformation. Although he has had multiple sclerotherapies, he has a persistent LM of his cheek. A gene panel to evaluate for pathogenic variants associated with overgrowth and vascular malformations was performed from LM tissue and identified somatic activating *PIK3CA* variant (c.3140A>T; p.H1047L). This variant has been reported before in lymphatic malformations and generalized lymphatic anomaly (RodriguezâLaguna et al 2019, Zenner et al 2019). Previous work in a mouse model demonstrated that expression of the *PIK3CA* p.H1047R variant in LECs causes lymphatic hyperplasia and dysfunction which is prevented by rapamycin, an mTOR inhibitor. To understand the etiology of our patient's variant and identify a targeted therapy, we performed functional analysis *in* vivo and *in vitro*. Expression of the variant in a 3D lymphatic organoid model using human dermal lymphatic endothelial cells resulted in more capillary-like structures emanating from spheroids *in vitro.* Expression of *pik3ca* p.H1047L in the lymphatic endothelium using a *mrc1a* promoter led to dilation of the thoracic duct and posterior cardinal veins in zebrafish. mTOR inhibition with rapamycin or OSI-027 was effective in rescuing the hyperproliferative sprouting phenotype observed *in vitro.* In conclusion, preclinical modeling of lymphangiogenesis recapitulated his disease and identified mTOR as a potential therapeutic target for the treatment of his condition. Future work is needed to identify an effective therapy *in vivo* and evaluate the efficacy of PI3K inhibitors or combined inhibitors of the PI3K-AKT-mTOR pathway.